Passeggia *et al. International Journal of STEM Education* (2026) 13:9
https://doi.org/10.1186/s40594-026-00599-x

International Journal of
STEM Education

RESEARCH Open Access

Exploring STEM identity profiles among undergraduate STEM students and their relationships with gender, discipline, persistence and engagement: a latent profile analysis

Raffaella Passeggia[1], Esposito Giovanna[1], Lucio Palazzo[2] and Italo Testa[3*]

Abstract

Background Research has thoroughly demonstrated that the intention to pursue and persist in a Science-Technology-Engineering-Mathematics (STEM)-related career is influenced by the extent to which students perceive themselves as a STEM person, namely their STEM identity. However, previous studies have mainly used variable-centered approaches to explore relationships between identity and its precursors, which might hide nontrivial differences amongst subgroups of a sample. This study examines profiles of undergraduate students based on their STEM identity, as defined by a multi-dimensional framework which includes perception of the self as a STEM person, perceived competence, interest, sense of belonging and perceived recognition. This study also investigates the role of students' gender, disciplinary area of the degree course, engagement in and intention to persist in the chosen undergraduate course in predicting profile membership. A convenience sample of 1114 STEM undergraduate students participated in this study (Mean age = 19.1 years; 50.9% female). An improved 6-factor scale was adopted to measure STEM identity. Profiles were extracted using a novel approach, Archetype Analysis, which allows to identify profiles that are more distinct and interpretable.

Results Archetypal Analysis revealed five distinct identity profiles: A1: Focused, A2: Relational; A3: Obligated; A4: Aspiring; A5: Detached. Female students were less likely to be classified within Archetypes Focused and Obligated, which are characterised by high levels of STEM self-identification. Students enrolled in engineering programmes were more likely to be associated with Archetypes Relational, Obligated and Aspiring, as opposed to students enrolled in technology courses. Technology students tended to cluster within the low-identity archetype Detached. Engagement was a significant predictor of membership of Archetypes Relational and Obligated, which are both characterised by strong social relatedness and recognition. Finally, the intention to persist in the chosen degree programme was a significant predictor of membership of archetypes Focused and Obligated.

*Correspondence:
Italo Testa
italo.testa@unina.it
Full list of author information is available at the end of the article

Conclusions This study contributes to the growing body of research on STEM identity by employing a novel statistical analysis to study this construct from a person-centred perspective. This approach allowed us to identify individual differences in students' patterns of interrelationship among the identity dimensions used.

Introduction

In recent years, higher education has played an increasingly pivotal role in providing the social and economic resources necessary for individual and collective advancement. Nevertheless, the proportion of graduates from tertiary education institutions across Europe remains below desirable levels, presenting a critical obstacle to sustained economic growth and innovation (Eurostat, 2025). This issue is particularly pronounced in Science, Technology, Engineering and Mathematics (STEM) disciplines, which are traditionally associated with rapid labour market integration and higher career rewards (Rocca & Quintano, 2024). Despite these advantages, STEM fields continue to experience low enrolment rates, exacerbated by the persistent underrepresentation of women. This limits access to STEM-related professional opportunities and reduces the utilisation of available human capital (European Commission, 2021, 2022). This underutilisation has significant implications for the broader economic benefits derived from research and innovation activities (Hanson et al., 2018).

An increasing amount of research has emphasised the significance of *STEM identity*—defined broadly as the self-identification process by which individuals see themselves in relation to the STEM field (Brickhouse et al., 2000) and as members of a STEM community (Chemers et al., 2011)—in shaping students' educational and career intentions (Calabrese Barton et al., 2013; Carlone & Johnson, 2007). Drawing on prior work about students' identity in STEM-related disciplines such as astronomy (Colantonio et al., 2021), physics (Hazari et al., 2010, 2020), chemistry (Guo et al., 2022), mathematics (Cribbs et al., 2015), computer science (Liong et al., 2023; Mahadeo et al., 2020), and engineering (Godwin et al., 2016; Li et al., 2021), STEM identity was operationalized for this study as a multidimensional construct which features the following dimensions: *perception of the self*, namely how students see themselves with respect to the STEM domain; *performance* or *perceived competence*, namely the belief about their own capabilities to apply acquired knowledge in performing STEM domain-specific tasks or to understand the content knowledge related to a STEM discipline; *interest*, namely the psychological state that allows a person to re-engage with a particular STEM content over time due to an intrinsic motivation or a perceived utility value of the content itself; *sense of belonging*, namely perceived relatedness with the community of peers, and *perceived recognition*, namely how an individual perceive how others, as teachers or instructors recognize them as a STEM person. Evidence suggests that a stronger STEM identity among high school students increases their awareness of the motivations for pursuing STEM undergraduate programmes (Liu et al., 2023; Lockhart et al., 2022) and predicts enrolment in STEM majors (Robinson et al., 2018; Vincent-Ruz & Schunn, 2018). At university level, strong disciplinary identification has been linked to persistence in STEM courses and sustained commitment to learning (Findley-Van Nostrand & Pollenz, 2017; Fisher et al., 2022). Furthermore, STEM identity has been associated with a greater sense of belonging (Jones et al., 2013; London et al., 2011) and better academic performance (Seyranian et al., 2018).

Given these findings, investigating STEM identity at undergraduate level provides valuable insights into the mechanisms that support academic success and retention in STEM disciplines. However, despite the growing interest in STEM identity among scholars, few studies have adopted a quantitative person-centered approach to identifying different STEM profiles among undergraduate STEM students. Moreover, there is a need to rigorously identify the extent to which membership in these profiles can be predicted by demographic variables, as gender, and contextual variables related to the attended degree course, as the chosen disciplinary area, the engagement in and the intention to persist in the chosen undergraduate course. Therefore, the following two broad research questions guided our study:

RQ1) What profiles can be identified with regard to undergraduate students' STEM identity?

RQ2) To what extent is membership in the identified profiles predicted by gender, disciplinary area of, engagement in and intention to persist in the chosen undergraduate course?

To address these questions, this study begins by describing the theoretical framework underlying the adopted definition of STEM identity, followed by a review of the literature examining its relationship with the demographic and chosen contextual variables. Additionally, we critically assess the methodological approaches employed in prior quantitative research about STEM identity, highlighting the limitations of variable-centered analyses in capturing subgroup-specific dynamics, and propose a person-centered approach to studying STEM identity in a more nuanced and effective way.

Background

STEM identity

In this study, we adopted multiple theoretical perspectives to define STEM identity. The reason for not relying on a single theory is to strengthen our conceptualization in order to study the multidimensional construct of student STEM identity in a more detailed way. First, we built on the Social Identity Theory (Tajfel, 1981), according to which STEM identity is a type of social identity, which indicates the extent to which individuals see themselves in terms of their membership and belonging in a social group (Kim et al., 2018) or of their positional designation and recognition within a given social context (Avraamidou, 2020; Burke & Stets, 2023). On such basis, an individual's STEM identity is hence the result of a dynamic process based on personal and social negotiations with others and the personal life environment (Cheryan et al., 2015; Gee, 2000; Kozoll & Osbourne, 2004). Second, we draw on Interest and Expectancy-Value Theories since previous studies suggest that STEM identity is influenced by motivational beliefs towards the specific STEM field (Atik et al., 2023; Chen et al., 2023a, b; Munoz et al., 2025; Stringer et al., 2020). Within these theoretical perspectives, STEM identity may develop also on the individuals' beliefs about the utility value they place on engaging with a STEM-related task, activity or career (Cohen et al., 2021) and can be sustained by external forms of support and potential reward (Eccles, 2009; Renninger & Hidi, 2021). Third, this study was guided by the Social Cognitive Career Theory (Bandura, 1997; Lent et al., 1994), given that STEM identity is influenced by the individual's perception of their own self-efficacy (Hazari et al., 2010, 2020; Munoz et al., 2025), and acts as a predictor of pursuing a career in the STEM domain (Hill et al., 2018; Sonnert et al., 2025). The adoption of these three theoretical perspectives accounts not only supports the multidimensionality of STEM identity in terms of *perception of the self*, *perceived recognition*, *sense of belonging*, *interest* and *perceived competence*, respectively, but also for the interconnectedness of these dimensions. For example, interest and utility value beliefs can be triggered by one's own personal negotiations with a STEM discipline (e.g., attitudes towards physics), social negotiations with STEM professionals (e.g. teachers or role models) and engagement in STEM-related experiences (e.g., extracurricular activities) (Kang et al., 2019; Tripp & Liu, 2024). Similarly, perceived competence, alongside personal interests, perceived barriers, and parental support (Pfeuffer et al., 2025), can contribute to the generation of outcome expectations and anticipated consequences that inform the decision to pursue a STEM career (Dou et al., 2019). Moreover, perceived recognition, which refers to judgements received both explicitly and implicitly from others in roles such as instructors or teachers is closely linked to social persuasion, which is one of the sources of self-efficacy (Britner & Pajares, 2006). A graphical representation of the adopted theoretical framework for STEM identity is reported in Fig. 1.

Factors influencing STEM identity

The role of gender. Most studies agree that girls report a significantly lower sense of identity in STEM subjects than boys (e.g. Hill et al., 2018). This suggests that a mismatch between how girls see themselves and how they see themselves as STEM students may contribute to the underrepresentation of women in STEM subjects (Hazari et al., 2020). In other words, girls who perceive a conflict between their STEM identity and their identity as a woman are less likely to pursue a STEM career (Miles & Naumann, 2021). This has been confirmed by other studies showing that identifying as a girl is negatively associated with participation in STEM degree programs, even when educational and demographic variables are taken into account (Godec et al., 2024). Similarly, the more that girls implicitly associate STEM with men or with the nerd-genius stereotype, the lower their STEM identification (Starr, 2018). Moreover, stereotype threat may cause girls to perceive higher social pressure in STEM fields (Kessels, 2005) or reduce their perception of the usefulness of science in helping others and society (Smith et al., 2015).

The role of student engagement. Student engagement has received increasing attention as a factor that can play an important role in sustaining students' persistence in the STEM field (Freda et al., 2023; Passeggia et al., 2023, 2025). Traditionally, engagement refers to a positive and proactive attitude towards a task or activity, capturing the quality of participation, investment and commitment (Appleton et al., 2006; Fredricks et al., 2004). Previous studies have suggested a relationship between identity and specific dimensions of academic engagement, such as self-regulated learning (Matthews et al., 2014). Furthermore, engagement with a particular course may predict students' identification with the institution and its values (Phan et al., 2016). Finally, engagement is rooted in social relations (Fredricks et al., 2019) and shared experiences that can influence future expectations and identification with the desired professional role (Shernoff et al., 2016). For these reasons, it is reasonable to expect that engagement in a STEM university course could predict STEM identity. However, these relationships have not yet been fully investigated.

The role of STEM disciplinary area. Literature suggests that STEM identity formation is not uniform across STEM areas (Dou et al., 2021). Rather, it reflects the cultural norms, practices, and expectations embedded within each discipline (Reinholz et al., 2019). For instance, Williams and George-Jackson (2014) found

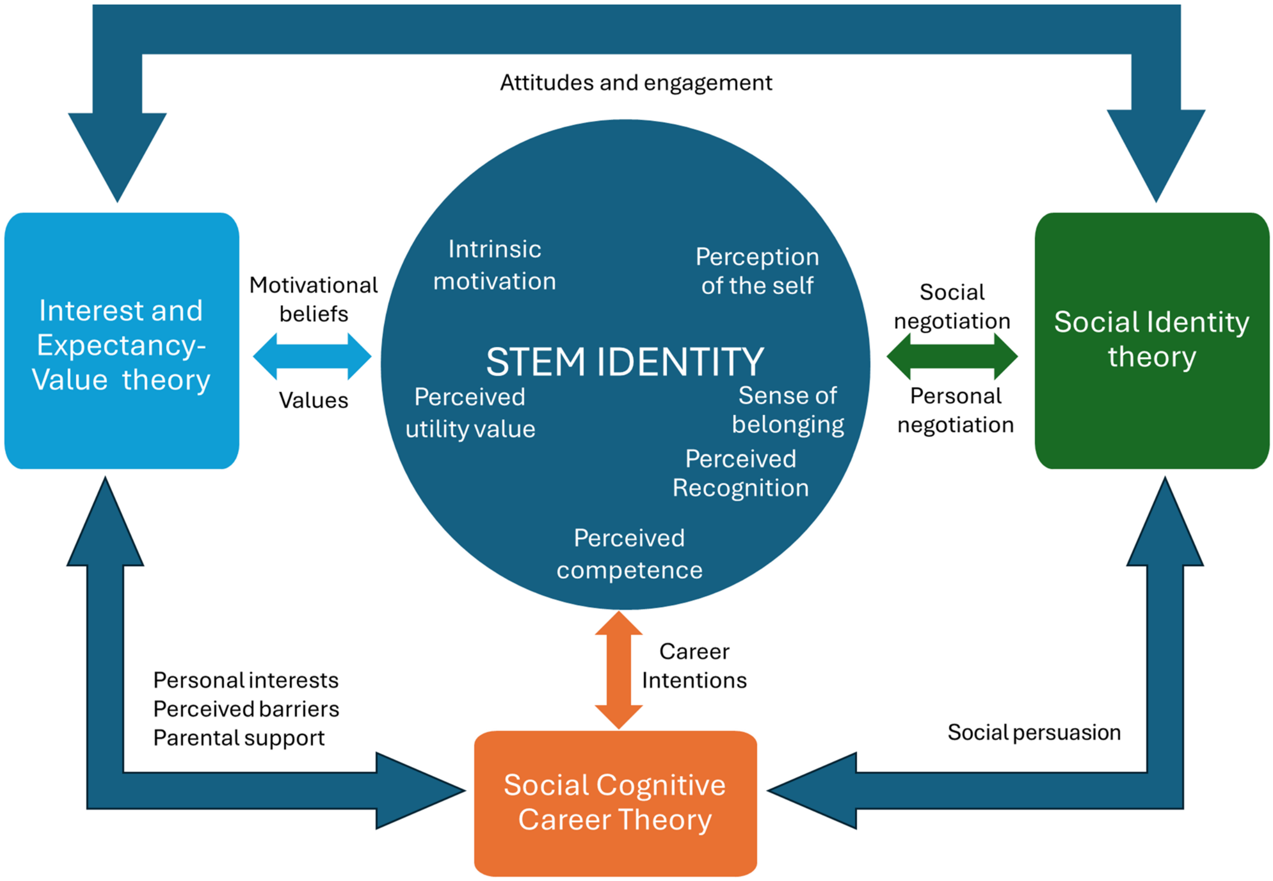

Fig. 1 Model for the STEM Identity framework

significant differences in self-reported identification as a scientist among 1881 undergraduates enrolled in STEM courses compared to those in agricultural and biological sciences or health sciences and psychology. Chen et al. (2023) demonstrated that high school students' science identities varied across physics, chemistry, and biology, with physics identity exerting the strongest influence on intentions to pursue science majors. In the study by Silva et al. (2025) about STEM identity across biology, chemistry, and mathematics among undergraduate students, findings revealed that students' major and prior informal STEM experiences significantly predicted disciplinary identity, with STEM majors reporting higher identity scores than non-STEM peers. This aligns with findings by Jiang et al. (2025), who indicate in their analysis that the strength of STEM identity's relationship with career intentions varies by identity type (e.g., physics vs. math identity), confirming that disciplinary differences are a key moderator of STEM identity. Overall, these studies highlight that STEM identity is not a monolithic construct across STEM-related disciplines, but it can be influenced by discipline-specific features. For these reasons, it is reasonable to expect that attending a specific STEM university degree course may reflect differences in the STEM identity.

The role of persistence. Differences in STEM identity are closely related to persistence in STEM undergraduate programs. Students who develop a strong STEM identity are more likely to remain engaged and complete their degrees, while those with weaker identities often leave STEM pathways (Patrick et al., 2018). Research highlights that some of dimensions included in our conceptualization of STEM identity, such as sense of belonging, perceived recognition and perceived competence, are related to persistence in STEM. For example, Estrada et al. (2019) found that social support networks through recognition and community significantly enhance graduate students' persistence in biomedical fields, which in turn reinforce disciplinary identity. Kuchynka et al. (2019) demonstrated that pre-college STEM engagement programs improve persistence by fostering perceived competence and sense of belonging early on. At the undergraduate level, Fisher et al. (2022) reported that students who left physics cited lack of identity alignment and insufficient support as key reasons for attrition, confirming results obtained by Patrick et al. (2018) in engineering field. Finally, Martin-Hansen (2018) showed that

STEM identity and persistence can reciprocally reinforce one another through meaningful support activities—such as mentoring. Overall, these studies underscore that students' persistence is not solely a function of their academic ability but can be related to how they perceived themselves as STEM persons.

Justification for the use of person-centered approach to study STEM identity

Quantitative studies on science and STEM identity have predominantly used variable-centered methodologies, such as structural equation modelling, to examine the links between identity constructs and their antecedents or contextual influences (Dou & Cian, 2022; Lin et al., 2025; Munoz et al., 2025; Starr et al., 2020). However, while these approaches have made significant contributions to STEM identity research, they are inherently limited in their ability to capture heterogeneity within populations. Specifically, these approaches assume homogeneity in relationships across individuals and infer population-level trends from average values within a sample (Morin et al., 2018), thereby overlooking potential variations among subgroups within the sample (Morin et al., 2011). Furthermore, the complexity of inter-variable relationships in these models can impede the derivation of clear, actionable insights for specific subpopulations. The common practice of identifying subgroups through mean or median splits further risks introducing sample-based bias (Mäkikangas et al., 2018). In contrast, person-centered approaches—aligned with an ecological systems perspective—conceptualize individuals as integrated psychological, biological, and social beings (Magnusson, 2003). These approaches emphasize identifying subgroups of individuals who exhibit similar patterns of attributes or inter-variable relationships within a given context (Harring & Hodis, 2016; Laursen & Hoff, 2006;). By capturing this heterogeneity, person-centered methods offer a more nuanced understanding of individual differences and developmental trajectories (Howard & Hoffman, 2018). Moreover, by simultaneously considering multiple variables, they provide a holistic perspective, which is particularly valuable for interpreting complex behaviors (Marsch et al., 2009). Importantly, person-centered analyses are often more applicable in educational contexts, as they support the design of tailored interventions and support strategies for diverse student populations (Litalien et al., 2019).

Person-centered methodologies provide an appealing alternative approach to examining STEM identity in quantitative studies. Specifically, these approaches enable researchers to classify individuals, such as undergraduates, based on their scores across multiple identity dimensions. These methods provide a distinct lens through which to examine how these dimensions interact within individuals rather than assuming uniform effects across a population, as is typical in structural equation modelling (Lee & Yoo, 2020). Furthermore, by identifying latent profiles, researchers can evaluate whether these groupings represent meaningful subpopulations and investigate the predictors of profile membership in relation to variables identified in the identity literature. Despite their potential, person-centered approaches have only recently been applied in quantitative research about STEM identity at university level. For instance, Robinson et al. (2019) employed latent growth modelling to analyze science identity trajectories among 1,669 students enrolled in an introductory chemistry course. They identified three distinct profiles—High-and-Stable (53.8%), Moderate-and-Slightly-Increasing (40.5%), and Moderate-and-Declining (5.6%)—revealing that, while most students' identities remained stable, some experienced significant changes within a single semester. Initial identity levels and stability were both significant predictors of STEM achievement and persistence. Finally, a recent study examined the development and stabilization of identity profiles among first-year engineering students (Lockhart & Rambo-Hernandez, 2024). The researchers identified three stable identity categories—Medium, Medium-High and High—and found that women were more likely than men to transition into the higher categories over time. Notably, the study also revealed that calculus readiness at the start of college significantly influenced persistence in engineering; students with a strong engineering identity but lacking calculus readiness were more likely to change their major or discontinue their studies. Overall, these results support the usefulness of person-centered approaches in identifying heterogeneity within large samples.

Methods

Participants

A convenience sample of $N = 1{,}114$ freshmen students attending 20 different STEM undergraduate degree courses (mean age = 19.1 years; 50.9% female) was involved in the study. About 28% attended an Engineering course, 32% a Science/Mathematics course, and 40% a Technology course. Engineering degree courses included: aerospace (5.5%), automation (2%), biomedical (5%), chemical (2%), civil (2%), electronics (2%), management (4.5%), mechanical (4%), others (1%). Science degree courses included: biology (13%), chemistry (7%), geology (1%), physics (2%), mathematics and statistics (4.5%), natural sciences (3.5%), others (1%). Technology degree courses included: architectural design (8%), biotechnology (1.5%), computer science and informatics (16%), pharmacy (8%), quality control (3.6%), technologies for animal productions (2%), others (1%). As expected, there was a significant association between gender and the three macro categories of degree courses ($\chi^2 = 22.435$, $d.o.f. = 2$, $p < 0.001$), with girls overrepresented in the science courses and boys overrepresented in the engineering courses.

The data were collected during in-presence lessons in the second semester of the first year. We remind that, in Italy, students enrolled in a STEM university course are required to follow only STEM-related classes, which typically include calculus, algebra, and at least one introductory class of biology, chemistry, informatics, physics, etc.… depending on the specific degree course. Similarly, students enrolled in a humanities-related university courses are not required to follow any STEM-related course. Therefore, in our sample, there were no students attending these courses for general education requirements. Moreover, all the degree courses were open, namely students were not required to pass any score in the entrance exam to attend the course. The score in the entrance exam is used only to determine if the students have to take first the calculus exam.

All participants provided written informed consent prior to participation and were fully briefed on the handling of their data. Specifically, students were informed that their responses would remain anonymous, that the data would be used exclusively for research purposes, and that all data would be stored in compliance with the General Data Protection Regulation (GDPR) 2016/679. The study was conducted in accordance with the ethical principles outlined in the 1964 Declaration of Helsinki and its subsequent amendments or equivalent ethical standards. Ethical approval for the research was granted by the Ethics Committee of the authors' university (Protocol PG2024/0011629, dated 29/01/2024).

Measures

For the present study, to measure the STEM identity dimensions in agreement with the adopted theoretical perspectives, we used six scales for a total of 26 items adapted from already existing instruments. Specifically, the following scales were used (see Table 1):

- *Perception of the self as a STEM person (PS)*, four-item scale adapted from Robinson et al., (2019). Example item: *I consider myself a STEM person*.
- *Perceived competence (PC)*, five-item scale adapted from Robinson et al., (2019). Example item: *I can complete even the most difficult tasks in STEM subjects if I try*.
- *Student' interest (SI)*, adapted from two subscales of the Academic Motivation Scale (AMS, Alivernini & Lucidi, 2008), one related to *intrinsic motivation (SI_I)* (4 items, example item: *I am interested in the things we do in the degree course I am enrolled in*), the other related to *perceived utility value (SI_U)* (4

Table 1 Description of the variables used in the study

Variable	Use in the study	Type	Number of items	Interval	Survey question	Survey options	Reference
Perception of the self as a STEM person (PS)	Profile indicator	Scale	4	1–5	Please indicate the extent to which you agree with the following statements	1 = strongly disagree 5 = strongly agree	Robinson et al. (2019)
Perceived Competence (PC)	Profile indicator	Scale	5	1–5	Please indicate the extent to which you agree with the following statements	1 = strongly disagree 5 = strongly agree	Robinson et al. (2019)
Student Interest: intrinsic motivation (SI_I)	Profile indicator	Scale	4	1–7	Please rate the extent to which each statement is true for you	1 = not at all 7 = completely	Alivernini & Lucidi, 2008
Student Interest: perceived utility (SI_U)	Profile indicator	Scale	4	1–7	Please rate the extent to which each statement is true for you	1 = not at all 7 = completely	Alivernini & Lucidi, 2008
Sense of belonging (SB)	Profile indicator	Scale	5	1–5	Please indicate the extent to which you agree with the following statements	1 = strongly disagree 5 = strongly agree	Freda et al., 2023
Perceived recognition (PR)	Profile indicator	Scale	4	1–5	Please indicate the extent to which you agree with the following statements	1 = strongly disagree 5 = strongly agree	Freda et al., 2023
Gender	Covariate	Categorical	1	–	Which of these BEST describes your gender at birth?	"Female" "Male"	–
Disciplinary area	Covariate	Categorical	1	–	Which degree course are you attending at the moment?	List of all STEM degree courses at the authors' university	–
Engagement in the chosen course	Covariate	Scale	16	1–5	Please indicate the extent to which you agree with the following statements	1 = strongly disagree 5 = strongly agree	Freda et al., 2023
Intention to persist in the chosen course (R)	Covariate	Scale	4	1–5	Please indicate the extent to which you agree with the following statements	1 = strongly disagree 5 = strongly agree	Freda et al., 2023

items, example item: *the STEM subjects in my course are important for what I have decided to do in the future*),

- *Sense of belonging (SB)*, a 5-item subscale adapted from an instrument previously validated by our group (Freda et al., 2023; example item: *I feel part of a group of friends in my STEM course*).
- *Perceived recognition (PR)*, a 4-item scale adapted from the same instrument (Freda et al., 2023; example item: *My STEM teachers consider me as a STEM person*).

A confirmatory factor analysis supported the validity of the 6-factor structure of our instrument: $\chi^2/d.o.f. = 3.631$, $p < 10^{-3}$, $RMSEA = 0.049$, $CFI = 0.97$, $TLI = 0.96$. All standardized factor loadings in the model were statistically significant and higher than the ideal value of 0.70 (Hair et al., 2009). Average Variance Extracted (AVE) was greater than the 0.5 threshold (Fornell & Larcker, 1981) for each dimension: 0.72, 0.65, 0.67, 0.71, 0.68, 0.60, respectively. Moreover, each AVE estimate was greater than the squared correlation estimates for any two dimensions (see Table 2). Overall, such values confirm convergent and discriminant validity of the instrument (Hair et al., 2009), namely that the factors are well separated and measure different dimensions of the STEM identity construct.

The factors influencing STEM identity reviewed in the background section were treated as covariates and measured as follows (see Table 1): (1) *gender at birth* was asked at the end of the survey and coded as a binary variable (1 = Female); (2) *disciplinary area* was coded as a categorical variable with three values–Engineering, Technology and Science/Mathematics–according to the categorization described above; (3) *engagement in the chosen course* and (4) *intention to persist in the chosen course* were measured through a 16-item and a 4-item reversed scale, respectively, both adapted from the same instrument validated by our group (Freda et al., 2023). Example items of the two scales are: *Going to my university degree course is a great opportunity for me* and *Sometimes I think about leaving university* (R), respectively.

Data analysis

To address the first research question (RQ1), we identified student profiles using Latent Profile Analysis (LPA) with Z-scores for each STEM identity dimension (self-identification, perceived competence, interest, sense of belonging, and recognition) as indicators. Profile extraction was performed using Archetypal Analysis (Cutler & Breiman, 1994), implemented in *R* with the *archetypes* package (Eugster & Leisch, 2009). Archetypal Analysis is an unsupervised learning technique that represents each observation as a convex combination of a set of archetypes—extreme points on the convex hull of the data (Seth & Eugster, 2016). Archetypal Analysis can be considered as a technique for dimensionality reduction aimed to summarize complex data into a set of more

Table 2 Descriptive statistics of measured variables in the study

Variable	1	2	3	4	5	6	7	8
1. Perception of the self as a STEM person (PS)	0.90 [a]							
2. Perceived competence (PC)	0.67**	0.91 [a]						
3. Student Interest: intrinsic motivation (SI_I)	0.44**	0.43**	0.86 [a];					
4. Student Interest: perceived utility (SI_U)	0.34**	0.33*	0.68**	0.91 [a]				
5. Sense of belonging (SB)	0.26**	0.27**	0.36**	0.30**	0.89 [a]			
6. Perceived recognition (PR)	0.26**	0.31**	0.36**	0.28**	0.31**	0.90 [a]		
7. Engagement in the chosen course	0.47**	0.47**	0.63**	0.63**	0.53**	0.43**	0.94[a]	
8. Intention to persist in the chosen course (R)	− 0.27	− 0.31**	− 0.33**	− 0.27**	− 0.14**	− 0.21**	− 0.38**	0.77 [a]
Average	3.58	3.47	5.49	5.68	3.80	3.46	3.92	1.87
Standard deviation	1.02	0.91	1.41	1.53	1.09	0.95	0.83	0.84
Min	1	1	1	1	1	1	1	1
Max	5	5	7	7	5	5	5	5
Asymmetry	− 0.495	− 0.554	− 1.114	− 1.252	− 0.926	− 0.493	− 1.086	1.209
Kurtosis	− 0.307	0.113	0.818	0.865	0.028	-0.121	1.299	1.209
Girls (N = 569)	3.50	3.32	5.43	5.64	3.91	3.38	3.99	1.79
Boys (N = 545)	3.66	3.63	5.54	5.71	3.68	3.54	3.84	1.95
t (d.o.f. = 1,112)	− 2.669**	− 5.696***	−1.256	0.792	3.455***	− 2.904**	3.059**	− 3.174***
Engineering (N = 309)	3.82	3.58	5,47	5,71	3,73	3,21	3,98	1,83
Science (N = 337)	3.57	3.46	5,46	5,65	3,92	3,68	3,88	1,86
Technology (N = 468)	3.43	3.41	5,51	5,67	3,76	3,47	3,90	1,90
Welch's F	14.118***	3.491*	0.165	0.117	3.532*	21.928***	1.080	0.534

[a] McDonald's omega; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

interpretable variables, the archetypes, similar to what happens in Principal Component Analysis (PCA). For instance, like in PCA, results of Archetypal Analysis can be represented in a set of principal axes or components. However, unlike the PCA, where the principal components extracted provide the directions of maximal variation, the number of principal axes and archetypes needed to represent the same features of a given data set can be different since Archetypal Analysis is more intrinsically related to data and less sensitive to outliers (Marinetti et al., 2007).

Formally, given a data matrix $X \in \mathbb{R}^{n\times p}$ the method seeks to identify k archetypes Z such that each data point x_i can be approximated as:

$$x_i \approx \sum_{j=1}^{k} \alpha_{ij} z_j$$

$$\text{where } \alpha_{ij} \geq 0 \text{ and } \sum_{j=1}^{k} \alpha_{ij} = 1.$$

For each data point x_i, the coefficients α_{ij}, with $j = 1, \ldots, k$ can be seen as a proximity score of each data point to their respective archetypes. Each archetype z_j itself can be seen as a convex combination of the original data points:

$$z_j = \sum_{i=1}^{n} \beta_{ji} x_i$$

$$\text{with } \beta_{ji} \geq 0 \text{ and } \sum_{i=1}^{n} \beta_{ji} = 1.$$

The combined values of the α_{ij} and β_{ij} are chosen to minimize the residual sum of squares (RSS):

$$RSS = \sum_{i=1}^{n} \left\| x_i - \sum_{j=1}^{k} \alpha_{ij} z_j \right\|^2$$

The estimation is performed via an alternating minimization algorithm, as implemented in the archetypes package, which iteratively updates the archetype and coefficient matrices to minimize the residual sum of squares between the observed data and their archetypal approximations. This two-level convexity ensures that archetypes are interpretable as "extreme" profiles within the data. Unlike centroid-based clustering algorithms (e.g., *k-means*), which minimize within-cluster variance and tend to produce clusters around the mean, Archetypal Analysis maximizes the separation between archetypes by focusing on the boundaries of the data distribution. This results in clusters (archetypal profiles) that are more distinct and interpretable, especially when the goal is to identify prototypical patterns rather than average trends (Bauckhage & Thurau, 2009). For such reasons, Archetypal Analysis has been successfully applied in various fields, including biology (Milite et al., 2025), climate science (Hannachi & Trendafilov, 2017), financial education (Palazzo et al., 2024), and economics (Abbasiharofteh & Hermans, 2025), demonstrating its versatility for uncovering latent structure in complex datasets. The optimal number of archetypes was determined using the elbow method by plotting the sum of residual squares against the number of archetypes and identifying the inflection point in the scree plot. To classify data points, we applied the nearest archetype rule (Ragozini et al., 2017), assigning each respondent x_i to the archetype z_j for which they had the highest α_{ij}. In the Supplemental Material, we report two additional cluster analyses, one based on hierarchical cluster analysis, the other one based on a different latent profile analysis algorithm.

To address the second research question (RQ2), we used multinomial logistic regression to examine the predictive value of gender, disciplinary area of the degree course, academic engagement, and persistence for the archetypal profile membership. This analysis was conducted in SPSS v. 30. Two models were estimated, Model 1 and Model 2. Model 1 included the main effects of the covariates, while Model 2 added interaction terms between gender and disciplinary area, given the association between gender and the three type of degree course. We remind that in both Model 1 and 2, gender was coded as a binary variable (female = 1), while degree disciplinary area was coded as a categorical variable (modalities = engineering, science, technology), from which three dummy variables were calculated and entered into the models. Model fit was evaluated using the likelihood ratio chi-square test, Pearson and deviance chi-square tests, and McFadden's pseudo R-square.

Results

Descriptive statistics of the measured variables

Table 2 shows the correlations between each of the measured variables, as well as the differences between genders and disciplinary areas. Descriptive analysis shows that all variables follow a normal distribution. All STEM identity dimensions are positively correlated with each other and with engagement in the chosen course. The reverse scale of intention to persist shows negative correlations with all STEM identity dimensions and engagement. Girls score significantly lower than boys in terms of their perception of themselves as STEM individuals, their perceived competence, and their perceived recognition. However, they score significantly higher than boys on the sense of belonging dimension. There are no significant gender differences in the interest subdimensions. There are statistically significant differences across the three categories of degree course for the perception of oneself as a STEM person, perceived competence, sense of belonging, and perceived recognition. Students on engineering courses have the highest scores on the self-identification with STEM and perceived competence scales, while students on science courses have the highest scores on the sense of belonging and perceived recognition scales. There are no significant differences across the three areas for the interest, engagement and persistence scales.

STEM identity profiles

Based on the RSS (residual sum of squares), the best model has 5 archetypes. The percentile plot of the profiles (Fig. 2) shows the structural properties of each emerging archetype.

Archetype 1 (A1) features archetypal students with high values in all STEM identity dimensions, except sense of belonging. About 11% of the students present the structural properties of this archetype. We called this archetype *Focused*. The main characteristic of the second archetype (A2) is the lack of self-identification and perceived competence in STEM. However, students associated with this archetype are highly interested in STEM and perceive themselves to be accepted and recognized by both peers and instructors. We called this archetype *Relational*. About 17.5% of the students are assigned to this archetype group. The majority of students (39%) of the sample present structural characteristics similar to Archetype 3 (A3), which features high values in self-identification, perceived competence and recognition, but very low values in interest. We called this archetype *Obligated.* Archetype 4 (A4) is similar to Archetype 1, except for the low values of perceived recognition. About 13% of the students are associated with this Archetype, which we called *Aspiring.* Finally, Archetype 5 (A5) is the archetypal student who does not identify with STEM in all its dimensions, with a low interest in STEM subjects (*Detached*). About 19% of the students belong to this archetype. The distribution of the archetypes plotted in the plane of the first two principal axes of the data is reported in Fig. 3.

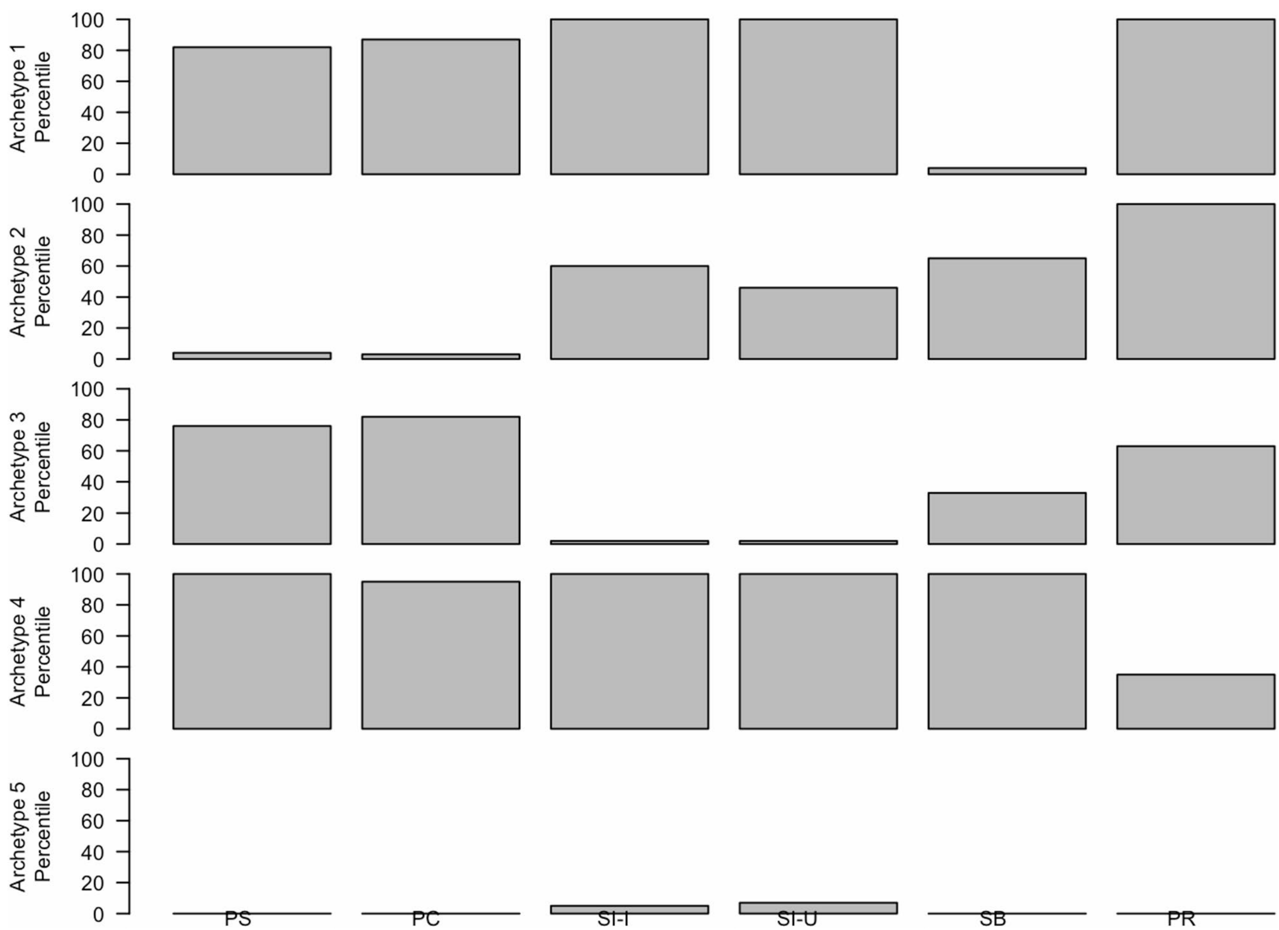

Fig. 2 Percentile plot of the profiles of STEM identity. *PS = Perception of the self as a STEM person; PC = Perceived Competence; SI-I = Students' Interest: Intrinsic; SI-U = Students' Interest: Utility; SB = Sense of Belonging; PR = Perceived Recognition. Archetype 1: Focused, Archetype 2: Relational; Archetype 3: Obligated; Archetype 4: Aspiring; Archetype 5: Detached*

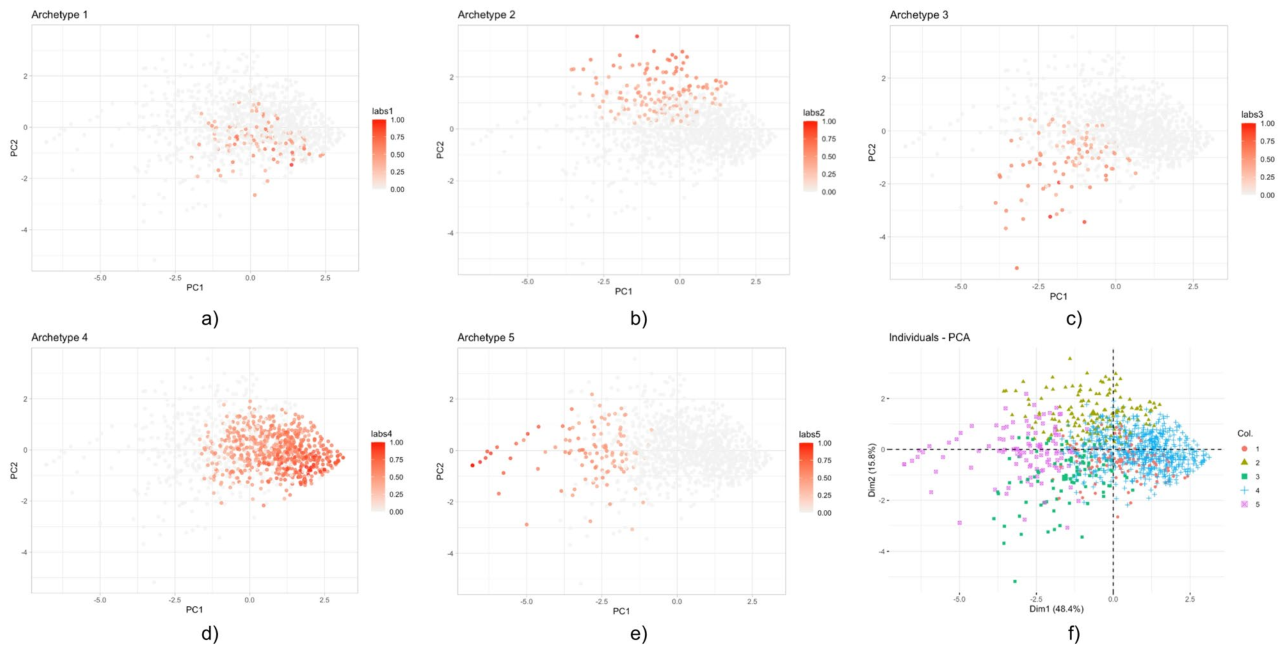

Fig. 3 Archetype distribution plotted in the Principal Components (PC) plane. **a** Archetype 1 (*Focused*); **b** Archetype 2 (*Relational*); **c** Archetype 3 (*Obligated*); **d** Archetype 4 (*Aspiring*); **e** Archetype 5 (*Detached*); **f** all archetypes. Variance explained by the two components is reported in frame f)

Table 3 Results from multinomial logic regression for profile membership–model 1. Reference profile = A5

	Archetype A1 versus A5		Archetype A2 versus A5		Archetype A3 versus A5		Archetype A4 versus A5	
	Coeff. (SE)	OR	Coeff. (SE)	OR	Coeff. (SE)	OR	Coeff. (SE)	OR
Intercept	0.61 (0.92)		– 1.73 (0.84)*		– 0.20 (0.71)		6.48 (0.91)***	
Engagement	0.05 (0.19)	1.05	0.44 (0.17)*	1.56	0.52 (0.15)***	1.67	– 2.06 (0.21)***	0.12
Persistence (R)	– 0.49 (0.16)**	0.61	– 0.18 (0.13)	0.83	– 0.44 (0.12)***	0.65	– 0.11 (0.15)	0.90
Gender (female)	– 1.15 (0.24)***	0.32	– 0.26 (0.22)	0.77	– 1.10 (0.18)***	0.36	0.04 (0.27)	1.04
Engineering	0.55 (0.31)	1.73	1.48 (0.26)***	4.40	0.59 (0.25)*	1.80	1.35 (0.33)***	3.87
Science	0.27 (0.26)	1.31	– 0.31 (0.26)	0.74	0.32 (0.19)	1.37	0.01 (0.30)	1.04

* *p* < .05. ** *p* < .01. *** *p* < .001. SE = Standard Error; OR = Odds Ratio. Archetypes: A1: Focused, A2: Relational; A3: Obligated; A4: Aspiring; A5: Detached

Table 4 Results from multinomial logic regression for profile membership–model 2. Reference profile = A5

	Archetype 1 versus 5		Archetype 2 versus 5		Archetype 3 versus 5		Archetype 4 versus 5	
	Coeff. (SE)	OR	Coeff. (SE)	OR	Coeff. (SE)	OR	Coeff. (SE)	OR
Intercept	0.70 (0.93)		– 1.50 (0.86)		– 0.15 (0.72)		6.72 (0.94)***	
Engagement	0.05 (0.19)	1.05	0.42 (0.17)*	1.52	0.51 (0.15)***	1.66	– 2.09 (0.21)***	0.12
Persistence (R)	– 0.50 (0.16)**	0.61	– 0.19 (0.13)	0.83	– 0.44 (0.12)***	0.64	– 0.12 (0.15)	0.89
Gender (female)	– 1.28 (0.35)***	0.28	– 0.46 (0.32)	0.63	– 1.11 (0.26)***	0.33	– 0.17 (0.40)	0.85
Engineering	0.24 (0.39)	1.27	0.74 (0.38)	2.10	0.27 (0.32)	1.31	0.67 (0.46)	1.95
Science	0.27 (0.42)	1.30	0.38 (0.43)*	1.46	0.48 (0.35)	1.62	0.31 (0.49)	1.36
Engineering x female	0.47 (0.69)	1.61	1.49 (0.55)**	4.42	0.64 (0.52)	1.90	1.54 (0.68)*	4.64
Science x female	0.14 (0.55)	1.15	– 1.12 (0.55)*	0.33	– 0.21 (0.42)	0.81	– 0.42 (0.63)	0.66

* *p* < .05. ** *p* < .01. *** *p* < .001. SE = Standard Error; OR = Odds Ratio. Archetypes: A1: Focused, A2: Relational; A3: Obligated; A4: Aspiring; A5: Detached

Predictors of identity profiles

Results from the two multinomial logistic regressions are reported in Tables 3 and 4. The two full multinomial models were comprised of four binary logit models where archetypes *Focused*, *Relational*, *Obligated*, *Aspiring* were subsequently compared the baseline archetype *Detached* comprised of students indicating least identification with STEM. Concerning Model 1, the likelihood ratio chi-square test was statistically significant, $\chi^2(20) = 510.243$, $p < .001$, indicating our model containing the full set of predictors fit the data significantly better than a null (or intercept-only) model. The Pearson chi-square, $\chi^2(4048) = 3885.544$, $p = 0.966$, and Deviance chi-square, $\chi^2(4048) = 2656.903$, $p = 1.000$, tests were both not statistically significant, suggesting the model can be considered a good fit to the data. McFadden's pseudo R-square for this model was 0.15, indicating a moderate improvement in the fit of our full model relative to a model containing no predictors. Next, we tested the omnibus effect of each of the predictors in the model, which allowed us to determine which contributed significantly to the overall model fit. The likelihood ratio tests were all statistically significant ($p < 0.001$).

Table 3 presents the regression coefficients, standard error and odds ratios for each of the four binary logit models that refers to Model 1. For the first one, Focused vs. Detached, both persistence (reversed) and gender emerged as significant predictors holding the remaining effects constant in the model. The negative slope and odds ratio indicate that a female student is 3 times more likely than a male student to be associated with the Detached profile with respect to the Focused profile. Conversely, a student with higher intended persistence is about 1.6 times more likely to be associated with the Focused profile than the Detached profile than a student with higher intention to drop-out. For the Relational vs. Detached model, the intercept is slightly significant and positive, meaning that the likelihood of being associated with Relational profile over Detached profile for an individual with average values of engagement and intention to persist is significantly greater than 0, independently of their gender or attended degree course. We also note that engagement and enrollment in an engineering course emerged as significant predictors for this binary model. The positive slopes for both variables and the respective odds ratio indicate that a student having a higher engagement is about 1.6 times more likely to be associated with the Relational profile rather than the Detached profile, while being enrolled in an engineering course leads to an increasing likelihood of being associated with the Relational profile rather than the Detached profile of a factor of 4.4. Concerning Obligated vs. Detached model, we found that all predictors are significant except being enrolled in a Science course. For instance, being a female student increases the likelihood of being associated with the Detached profile vs. the Obligated profile of a factor of 2.8, while students enrolled in an engineering course are almost 2 times more likely to be associated with the Obligated profile than the Detached profile. Finally, for the Aspiring vs. Detached model, we note that the intercept in this binary logit model is highly significant and positive, thus individuals are in general more likely to be associated with the Aspiring profile than the Detached profile. However, while being a student in an engineering course further increases the likelihood of being associated with the Aspiring profile of a factor of about 4, students that have a higher engagement are 8 times more likely to be associated with the Detached profile.

Concerning Model 2, the likelihood ratio chi-square test was statistically significant, $\chi^2(28) = 535.84$, $p < .001$, indicating that our model contain the full set of predictors fit the data significantly better than a null (or intercept-only) model. The Pearson chi-square, $\chi^2(4040) = 3883.560$, $p = 0.990$, and Deviance chi-square, $\chi^2(4040) = 2631.305$, $p = 1.000$, tests were both not statistically significant, suggesting the model can be considered a good fit to the data. McFadden's pseudo R-square increased to 0.16, indicating however also in this case a moderate improvement in the fit of our full model relative to a model containing no predictors. Next, we tested the omnibus effect of each of the predictors in the model. The likelihood ratio tests were all statistically significant ($p < 0.001$), except for the course variable alone, whose contribution is not statistically significant ($p = 0.464$). Table 4 contains the regression coefficients, stander error and odds ratios for each of the four binary logit models that refer to Model 2. In the first model, Focused vs. Detached, the interaction terms are non-statistically significant, while both persistence (reversed) and gender remain as significant predictors. We note that a female student is now about 3.6 times more likely than a male student to be associated with the Detached profile with respect to Focused profile. For the Relational vs. Detached model, engagement is still a significant predictor, but we note that interaction terms are also significant. Specifically, while a female student in an engineering course is more likely of being associated with the Relational profile of a factor of 4.4, a female student in a science course is 3 times more likely to be associated with the Detached profile. The Obligated vs. Detached model is very similar to Model 1, as interaction terms are in this case not statically significant. Finally, for the Aspiring vs. Detached model, the intercept remains significantly different from zero, meaning that individuals with average values of engagement and intention to persist, would have a greater propensity to be associated in Aspiring profile than the Detached profile independently of gender and attended degree course. Engagement is still a significant predictor as for Model 1, while the odds ratio for the interaction term *engineering x gender* indicates that being a female student in an engineering course further increases the likelihood of being associated with Aspiring profile with respect to the Detached profile of a factor of about 4.6.

Discussion

This study contributes to the literature on STEM identity by using a person-centered quantitative approach to examine first-year undergraduate students enrolled in science, technology and engineering degree programs. Notably, this research is the first to utilize Archetype Analysis within the broader framework of latent profile analysis in this context. In response to the increasing focus on identity constructs in STEM education, we examined the number and characteristics of different STEM identity profiles, as well as how contextual factors such as gender, course type, engagement and intention to persist influence archetype membership.

RQ1

Our findings suggest that the diversity of STEM identity among first-year undergraduates can be best captured by five archetypes: A1: Focused, A2: Relational; A3: Obligated; A4: Aspiring; A5: Detached. The emergence of these five distinct profiles demonstrates that students in various STEM courses are not a homogeneous group but can instead be categorised into subgroups with unique identity characteristics. This result extends previous research, which has generally focused on the identity of scientists or engineers in isolation (Robinson et al., 2018, 2019; Lockhart & Rambo-Hernandez, 2024), by emphasising the nuanced and multifaceted nature of STEM identity across different fields of study (Silva et al., 2025).

Importantly, our analysis revealed more differentiated and internally coherent profiles than earlier studies about science and STEM-related identity at undergraduate level carried out by Robinson et al. (2019) and by Lockhart and Rambo-Hernandez (2024). In both studies, the emerging profiles reflected scores modulated around the means, with an unclear distinction in the indicator variables, for instance, between medium high and high identity profiles (Lockhart & Rambo-Hernandez, 2024) or between moderate and high identity profiles (Robinson et al., 2019). Differently, through Archetypal Analysis we were able to clearly identify the contribution of each of the chosen indicator variables—the STEM identity dimensions—thus allowing us to better differentiate the latent profiles. For instance, the most prevalent archetype (Obligated, 38%) corresponds to a student who may have pursued a STEM degree due to external validation (e.g. perceived recognition by peers or instructors) and a more internalized perception of the self, rather than intrinsic motivation. These students may therefore feel a sense of duty or expectation to persist in STEM but lack genuine interest. This result suggests that enrolment decisions in STEM may also be shaped by external social factors or perceptions of the utility of these courses for future careers (Dou & Cian, 2022), with students having little awareness of the specific content addressed in these courses (McComas & Burgin, 2020). Conversely, archetype Relational (17.5%) is typified by a genuine interest and intrinsic motivation in STEM subjects, a high perceived utility value, and external validation. However, it also indicates an uninternalized perception of the self and perceived competence. Students associated with this archetype are likely to be on the right track, but lack self-assurance, suggesting that not all students embark on their academic journey with a well-established STEM identity (Starr et al., 2020). Only about 25% of the sample is more aligned with archetypes Focused and Aspiring, which share high perception of the self, interest, due to both intrinsic motivation and utility value, and perceived competence, but differ in terms of sense of belonging and perceived recognition. This new finding suggests that even students who are committed to STEM and fully aligned with it may experience low recognition from peers and instructors cognitively and motivationally. This aligns with recent findings that external recognition may develop and become internalized differently from self-identification (Simpson et al., 2025; Sundstrom & Holmes, 2025). Finally, a concerning finding is that the second most common profile is associated with archetype Detached (19%), characterized by low scores in all STEM identity dimensions. Students who are more aligned with this archetype may be at a higher risk of dropping out (Kocsis & Molnár, 2025) and may experience exclusion or alienation. These factors are known to impact retention, particularly among underrepresented groups in specific STEM fields (London et al., 2011). This result may complement recent findings from a systematic review (Jiang et al., 2025) according to which STEM identity accounts for only 14% of the variance in STEM career intention. Therefore, our study concurs that future research should explore the role also of other cognitive and contextual factors in shaping a STEM career intention.

RQ2

With regard to our second research question, the present study provides further evidence that gender significantly influences STEM identity profiles among first-year undergraduates. However, our person-centred approach revealed different patterns within the sample. Specifically, female students were less likely to be classified within archetypes Focused and Obligated, which are characterised by high levels of perception of the self as a STEM person, perceived competence and recognition, compared to the low-identity archetype Detached, even when interactions between gender and disciplinary area were controlled for. This suggests that these dimensions of STEM identity may be particularly susceptible to gender disparities, in line with previous research in science (Alhadabi, 2021; Marshman et al., 2018; Miles & Naumann, 2021) and physics (Marshman et al., 2018). However, we also found that gender did not significantly predict membership of archetype Aspiring, characterised by high perception of the self as a STEM person, interest, and perceived competence, but low perceived recognition. This result therefore challenges the findings of previous studies in the physics domain, which suggest that even women majoring in fields in which they are not underrepresented (e.g. bioscience) may have a weaker sense of identity than their male counterparts due to a lack of recognition (Cwik & Singh, 2022). One possible explanation for this mixed evidence is the broader disciplinary representation in our sample, which included students from a great variety of STEM degree courses. Thus, different patterns in the identity dimensions may emerge in less homogeneous STEM samples, as suggested by a recent longitudinal study in the vocational research field (Gao et al., 2025). Furthermore, our results may reflect the fact that female students develop their STEM identity later in their academic careers, as suggested by studies in engineering contexts (Lockhart & Rambo-Hernandez, 2024).

Another contribution of this study is the finding that students enrolled in engineering programmes were more likely to be associated with Archetypes Relational, Obligated and Aspiring than with Detached, as opposed to those enrolled in technology courses. Notably, technology students tended to cluster within the low-identity archetype. This highlights a potential area of concern regarding persistence and dropout risk in these fields, given the established link between identity and academic retention (Jaiswal et al., 2022). Engineering students were almost four times more likely to belong to the archetypes Relational and Aspiring, which are characterised by high interest and relatedness. This finding supports previous research suggesting that engineering identity is strongly influenced by interest in and engagement with disciplinary practices (Bahnson et al., 2023; Godwin, 2016; Verdín, 2021). This trend was particularly evident among female engineering students, who were four times more likely than their male counterparts to be associated with these archetypes. This result aligns with previous research on the development of engineering identity among women (Godwin & Potvin, 2017; Lockhart & Rambo-Hernandez, 2024) and suggests that interest and social relatedness are particularly salient components in the identity formation process for female students. Likewise, when considering the roles of engagement and intention to persist, different patterns emerged within the sample. Engagement was a significant predictor of membership of archetypes Relational and Obligated, which are both characterized by strong social relatedness and recognition. This supports the idea that positive interactions with peers and instructors can have a beneficial effect on STEM identity, even when perception of the self as a STEM person has not yet been internalized or when both intrinsic motivation in and utility value of the subject are low (Shernoff et al., 2016; Zhoc et al., 2019). However, higher engagement was associated with a lower likelihood of belonging to the Aspiring archetype, which is characterized by high values across all identity dimensions except for perceived recognition. It should be noted, however, that, for the whole sample, being associated with archetype Aspiring versus Detached is significantly more likely, independently of all the other measured covariates, and that higher engagement decreases such likelihood while remaining significant. One possible interpretation is that perceived recognition from instructors constitutes a significant part of engagement in the degree course. Overall, these results suggest that engagement and STEM identity largely overlap in terms of recognition and sense of belonging.

Consistently with previous STEM field research (Guo et al., 2022; Lockhart et al., 2022; Vincent-Ruz & Schunn, 2018), the intention to persist in the chosen degree program was a significant predictor of membership of archetypes characterized by high perception of the self as a STEM person, perceived competence and recognition (Focused and Obligated versus Detached). However, we note that the students that tend to the Obligated archetype are likely to show low levels of intrinsic motivation in and perceived utility value of the contents studied in the chosen course. Thus, our results are also consistent with reviews that highlight the debated role of interest in fostering greater persistence and retention in higher education (Kocsis & Molnár, 2025; Li & Wong, 2019).

Finally, we note that our Archetype Analysis can significantly complement qualitative research on STEM identity (e.g. Matusovich et al., 2010; Simpson et al., 2025; Tripp & Liu, 2024). For instance, the differences between the extracted profiles can be contextualized by qualitative interviews with typical students of each profile or triangulated with qualitative narratives of their lived experiences to better understand the mechanisms underlying these differences. This was done for instance by Liong et al. (2023), who complemented their quantitative study with 821 female students aimed at uncovering profiles in computer science with structured interviews with 20 students aimed at identifying further insights into females' experiences with computer science. Overall, integrating the qualitative and quantitative person-centered approaches through mixed-methods designs enables researchers to triangulate findings, enhancing validity and offering a holistic understanding of STEM identity formation that captures both macro-level patterns and micro-level meanings (Creswell & Clark, 2017).

Limitations

Despite the novel approach adopted in this study to investigate STEM identity, several limitations should be acknowledged. Firstly, as is common with person-centered methodologies, the latent profiles identified are specific to our sample, which may limit the generalisability of these findings. Data were collected from students attending classes in person during the second semester of their first year, thereby excluding those who had withdrawn after the first semester or were not attending classes in person. Although efforts were made to mitigate sampling bias by including participants from 20 undergraduate degree programs across three STEM disciplines (science/mathematics, engineering and technology), the representativeness of the sample remains a concern. Future research should explore whether similar profiles emerge in different institutional contexts and at various stages of course attendance. Secondly, while the study examined key components of STEM identity, such as self-identification, perceived competence, interest, recognition and sense of belonging, this was done solely within an academic context. Other ways in which individuals may perceive themselves as part of STEM outside of academia were not considered. Thirdly, gender was operationalised as a binary variable. While this approach aligns with much of the existing literature and facilitated robust statistical analysis, it does not capture the full spectrum of gender diversity. Future studies should incorporate more inclusive measures of gender to better reflect the diversity of student experiences in STEM. Fourth, we accounted only for differences due to the different disciplinary STEM macro-areas, engineering, math and science, and technology. However, within the same disciplinary area, there could be potential variations that can occur due to the different instructors, classroom environment, and difficulty of content addressed. Future research studies may address such weakness by including the ones reported above as well as other contextual variables more related to the teaching practice of the degree courses.

Finally, academic performance data could not be collected for all participants. Consequently, the relationship between the identified profiles and academic outcomes could not be fully explored. Subsequent research should therefore aim to address this gap by examining how these profiles relate to performance at various stages of the academic career.

Conclusions

This study contributes to the growing body of research on STEM identity by employing a novel statistical analysis to study this construct from a person-centered perspective. This approach allowed us to identify individual differences in students' patterns of interrelationship among the identity dimensions used. Our findings revealed five distinct STEM identity profiles among first-year STEM course students. Notably, both gender and the specific course attended were found to significantly influence profile membership. Furthermore, the interaction between gender and degree course suggests that the internalisation of STEM identity may occur differently for female students than for their male counterparts. The results also indicate that STEM identity is shaped by students' engagement and their intention to persist with their chosen degree program, thereby reinforcing the critical role of engagement in academic trajectories. Future studies could fruitfully explore the origins and implications of these distinct profiles, as well as the interplay between STEM identity and other affective variables. The aim would be to elucidate how these relationships impact students' persistence and performance in STEM academic careers. Furthermore, the nuanced differences in identity profiles revealed by person-centered approaches, such as the Archetype Analysis used in the present study, suggest that more sophisticated STEM identity assessment tools can help to identify students at risk and inform the design of more effective, personalized support strategies.

The findings of this study have several important implications for STEM education at undergraduate level. Firstly, the persistent gender disparities in STEM identity, as evidenced by our study—particularly with regard to self-identification, perceived competence and recognition by instructors—highlight the need for targeted interventions to support female students, particularly in the fields of science and technology. Potential educational programs at undergraduate level aimed to develop STEM identity should prioritize strategies that foster these identity dimensions, such as mentorship opportunities, visible role models and inclusive classroom practices that actively recognize and validate female students' contributions (Miles & Vela, 2025; Singer et al., 2020). Secondly, in agreement with recent studies (Sonnert et al., 2025) variation in identity profiles across different STEM disciplines suggests that such programs should not adopt a one-size-fits-all approach but consider the unique challenges and motivations within each STEM discipline in order to mitigate also the drop-out risk and enhance persistence (Elsdon et al., 2025). For instance, tailored interventions that emphasize the significance of the habitus, i.e. personal dispositions that allow a person to act, think and orient themselves in the social world, technical-science capital, i.e. the set of knowledge and resources related to technical science, and social practice (Engstrom & Blom, 2024) may potentially enhance students' perception of the self as a STEM person, especially in disciplinary areas as Technology, where students were more likely to exhibit lower STEM identity and higher intention to drop-out from the degree course.

Thirdly, the finding that engagement predicts membership of identity profiles characterized by social relatedness and recognition, particularly among engineering students, suggests that creating opportunities for meaningful interaction with peers and instructors can have a positive influence on students' STEM identity. Institutions might therefore consider integrating active learning, group work and faculty-student interaction into STEM courses, in order to promote sense of belonging and recognition among freshmen students. Innovative curricula that emphasize collaborative projects, real-world applications, and community-building activities may be especially effective in strengthening STEM identity, particularly among underrepresented and underserved students (Barajas-Salazar et al., 2025; Richter & Kjellgren, 2024). Fourthly, the correlation between a strong sense of identity in STEM subjects and the intention to continue studying in this field highlights the importance of identity-focused interventions in improving retention rates. Programs that help students to articulate and develop their STEM identities, for example through reflective activities, career exploration and recognition of achievements, could be a valuable addition to well-established university practices. These programs may contribute to higher persistence rates, even among students who are initially uninterested or feel a low sense of relatedness (Graham et al., 2014; Thompson et al., 2023). In summary, these findings highlight the importance of multifaceted, context-sensitive and identity-aware educational practices in both formal in-school and informal out-of-school time STEM activities (e.g. Ovid et al., 2023; Sonnert et al., 2025). Addressing the various ways in which students experience and internalize their STEM identity can enable educators to provide more effective support for persistence, engagement, and success in all STEM disciplines.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40594-026-00599-x.

Supplementary Material 1.

Acknowledgements

We would like to thank the anonymous reviewers for their valuable feedback and suggestions which contributed to the improvement of this paper.

Author contributions

I.T. and G.E. designed the study. I.T. wrote the main manuscript text. I.T., R.P and L.P. analyzed the data. I.T. prepared figure 1. L.P. prepared the Figs. 2 and 3. All authors reviewed the manuscript.

Funding

The study was funded by Research Fund of University of Naples Federico II, fund n. 000012–ALTRI_DR_3429_FRA_Linea_B_ORIENTATION.

Data availability

Privacy restrictions apply to the availability of the data.

Declarations

Ethics approval and consent to participate

The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. The research project was approved by the Ethical Committee of University of Naples Federico II (prot. PG2024/0011629 of 29/01/2024).

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Humanities, University of Naples Federico II, Naples, Italy
[2]Department of Social Sciences and Humanities, University of Naples - L 'Orientale, Naples, Italy
[3]Department of Physics "Ettore Pancini", University of Naples Federico II, Naples, Italy

Received: 7 October 2025 / Accepted: 26 January 2026
Published online: 08 February 2026

References

Abbasiharofteh, M., & Hermans, F. (2025). Using archetypal analysis to derive a typology of knowledge networks in European bioclusters. *Regional Studies*, *59*(1), 2430354. https://doi.org/10.1080/00343404.2024.2430354

Alhadabi, A. (2021). Individual and contextual effects on science identity among American ninth-grade students (HSLS:09): Hierarchical linear modeling. *Research in Science & Technological Education*, *41*(3), 886–905. https://doi.org/10.1080/02635143.2021.1972959

Alivernini, F., & Lucidi, F. (2008). The academic motivation scale (AMS): Factorial structure, invariance and validity in the Italian context. *Testing Psychometrics Methodology in Applied Psychology*, *15*(4), 211–220. https://doi.org/10.4473/TPM.15.4.3

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology*, *44*(5), 427–445. https://doi.org/10.1016/j.jsp.2006.04.002

Atik, A. D., Sari, H. I., & Doğan, Y. (2023). The effects of closeness to nature, connectedness to nature and eco-friendly behaviours on environmental identity: A study of public university students in South-eastern Turkey. *Australian Journal of Environmental Education*, *39*(1), 80–94. https://doi.org/10.1017/aee.2022.20

Avraamidou, L. (2020). Science identity as a landscape of becoming: Rethinking recognition and emotions through an intersectionality lens. *Cultural Studies of Scince Education*, *15*, 323–345. https://doi.org/10.1007/s11422-019-09954-7

Bahnson, M., Satterfield, D., Perkins, H., Parker, M., Tsugawa, M., Cass, C., & Kirn, A. (2023). Engineer identity and degree completion intentions in doctoral study. *Journal of Engineering Education*, *112*(2), 445–461. https://doi.org/10.1002/jee.20516

Bandura, A. (1997). *Self-efficacy: The exercises of control*. W.H. Freeman and Company.

Barajas-Salazar, B. E., Almeida, M., Aguirre-Muñoz, Z., & Viveros, M. (2025). Culturally relevant informal STEM learning for underserved students: Effects of repeated exposure to the engineering design process. *Frontiers in Education*, *10*, 1534452. https://doi.org/10.3389/feduc.2025.1534452

Bauckhage, C., & Thurau, C. (2009). Making archetypal analysis practical. In J. Denzler, G. Notni, & H. Süße (Eds.), *Pattern Recognition, lecture notes in computer science* (Vol. 5748, pp. 272–281). Springer.

Brickhouse, N. W., Lowery, P., & Schultz, K. (2000). What kind of a Girl does science? The construction of school science identities. *Journal of Research in Science Teaching*, *37*(5), 441–458.

Britner, S. L., & Pajares, F. (2006). Sources of science Self-Efficacy beliefs of middle school students. *Journal of Research in Science Teaching*, *43*(5), 485–499. https://doi.org/10.1002/tea.20131

Burke, P. J., & Stets, J. E. (2023). *Identity theory* (2nd ed.). Oxford University Press.

Calabrese Barton, A., Kang, H., Tan, E., O'Neill, T. B., Bautista-Guerra, J., & Brecklin, C. (2013). Crafting a future in science: Tracing middle school girls' identity work over time and space. *American Educational Research Journal*, *50*(1), 37–75. https://doi.org/10.3102/0002831212458142

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching*, *44*(8), 1187–1218.

Chemers, M. M., Zurbriggen, E. L., Syed, M., Goza, B. K., & Bearman, S. (2011). The role of efficacy and identity in science career commitment among underrepresented minority students: Efficacy and identity in science career commitment. *Journal of Social Issues*, *67*(3), 469–491. https://doi.org/10.1111/j.1540-4560.2011.01710.x

Chen, S., Wei, B., & Zhang, H. (2023a). Exploring high school students' disciplinary science identities and their differences. *International Journal of Science and Mathematics Education*, *21*(2), 377–394. https://doi.org/10.1007/s10763-022-10257-7

Chen, J., Hasan, M. A., Du, X., & Kolmos, A. (2023b). Exploring students' perception of the influence of PBL elements on the development of engineering identity. *IEEE Transactions on Education*, *66*(4), 393–403. https://doi.org/10.1109/TE.2023.3258548

Cheryan, S., Master, A., & Meltzoff, A. N. (2015). Cultural stereotypes as gatekeepers: Increasing girls' interest in computer science and engineering by diversifying stereotypes. *Frontiers in Psychology*, *6*, 1–8. https://doi.org/10.3389/fpsyg.2015.00049

Cohen, S. M., Hazari, Z., Mahadeo, J., Sonnert, G., & Sadler, P. M. (2021). Examining the effect of early STEM experiences as a form of STEM capital and identity capital on STEM identity: A gender study. *Science Education*, *105*, 1126–1150. https://doi.org/10.1002/sce.21670

Colantonio, A., Marzoli, I., Puddu, E., Bardelli, S., Fulco, M. T., Galano, S., Terranegra, L., & Testa, I. (2021). Describing astronomy identity of upper primary and middle school students through structural equation modeling. *Physical Review Physics Education Research*, *17*, 010139.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage.

Cribbs, J. D., Hazari, Z., Sonnert, G., & Sadler, P. M. (2015). Establishing an explanatory model for mathematics identity. *Child Development*, *86*(4), 1048–1062. https://doi.org/10.1111/cdev.12363

Cutler, A., & Breiman, L. (1994). Archetypal analysis. *Technometrics*, *36*(4), 338–347. https://doi.org/10.1080/00401706.1994.10485840

Cwik, S., & Singh, C. (2022). Self-efficacy and perceived recognition by peers, instructors, and teaching assistants in physics predict bioscience majors' science identity. *PLOS ONE*, *17*(9), e0273621. https://doi.org/10.1371/journal.pone.0273621

Dou, R., & Cian, H. (2022). Constructing STEM identity: An expanded structural model for STEM identity research. *Journal of Research in Science Teaching*, *59*(3), 458–490. https://doi.org/10.1002/tea.21734

Dou, R., Hazari, Z., Dabney, K., Sonnert, G., & Sadler, P. (2019). Early informal STEM experiences and STEM identity: The importance of talking science. *Science Education*, *103*(3), 623–637. https://doi.org/10.1002/sce.21499

Dou, R., Cian, H., & Espinosa-Suarez, V. (2021). Undergraduate STEM majors on and off the pre-med/health track: A STEM identity perspective. *CBE—Life Sciences Education*, *20*(2). https://doi.org/10.1187/cbe.20-12-0281

Eccles, J. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist*, *44*, 78–79. https://doi.org/10.1080/00461520902832368

Elsdon, M., Hayman, R., & Agrawal, P. (2025). The impact of group-based experiential learning in developing undergraduate societal engineering skills: A case study within a united Kingdom Post-92 university. *European Journal of Engineering Education*, 1–17. https://doi.org/10.1080/03043797.2025.2552301

Engström, S., Blom, J. (2024) Technical science capital in relation to how students manage within higher technical education. *International Journal of Technology and Design Education,* 34, 1419–1439 https://doi.org/10.1007/s10798-023-09870-6

Estrada, M., Zhi, Q., Nwankwo, E., & Gershon, R. (2019). The influence of social supports on graduate student persistence in biomedical fields. *CBE—Life Sciences Education*, *18*(3). https://doi.org/10.1187/cbe.19-01-0029

Eugster, M. J. A., & Leisch, F. (2009). From Spider-Man to hero–Archetypal analysis in R. *Journal of Statistical Software*, *30*(8), 1–23.

European Commission. (2021). *She Figs. 2021: Gender in research and innovation: Statistics and indicators*. Publication office of the European Union. https://data.europa.eu/doi/ https://doi.org/10.2777/06090

European Commission. (2022). *2022 report on gender equality in the EU*. Publication office of the European Union. https://doi.org/10.2838/94579

Eurostat (2025). *Tertiary education statistics*. Retrieved on 20th September 2025 from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tertiary_education_statistics

Findley-Van Nostrand, D., & Pollenz, R. S. (2017). Evaluating psychosocial mechanisms underlying STEM persistence in undergraduates: Evidence of impact from a six-day pre–college engagement STEM academy program. *CBE—Life Sciences Education*, *16*(2). https://doi.org/10.1187/cbe.16-10-0294

Fisher, C. R., Brookes, R. H., & Thompson, C. D. (2022). I don't study physics anymore': A cross-institutional Australian study on factors impacting the persistence of undergraduate science students. *Research in Science Education*, *52*(5), 1565–1581. https://doi.org/10.1007/s11165-021-09995-5

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*, 39–50.

Freda, M. F., De Luca Picione, R., Esposito, G., Ragozini, G., & Testa, I. (2023). A new measure for the assessment of the university engagement: The SInAPSi academic engagement scale (SAES). *Current Psychology*, *42*(12), 9674–9690. https://doi.org/10.1007/s12144-021-02189-2

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Fredricks, J. A., Reschly, A. L., & Christenson, S. L. (2019). Interventions for student engagement: Overview and state of the field. In J. A. Fredricks, A. L. Reschly, & S. L. Christenson (Eds.), Handbook of student engagement interventions: Working with disengaged students (pp. 1–11). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-813413-9.00001-2

Gao, Y., Eccles, J. S., & Dicke, A. L. (2025). Not a pipeline but a highway: Men's and women's STEM career trajectories from age 13 to 25. *Journal of Vocational Behavior*, *156*, 104067. https://doi.org/10.1016/j.jvb.2024.104067

Gee, J. (2000). Identity as an analytic lens for research in education. *Review of Research in Education*, *25*, 99–125.

Godec, S., Archer, L., Moote, J., Watson, E., DeWitt, J., Henderson, M., & Francis, B. (2024). A missing piece of the puzzle? Exploring whether science capital and STEM identity are *associated* with STEM study at university. *International Journal of Science & Mathematics Education*, *22*, 7, 1615–1636. https://doi.org/10.1007/s10763-023-10438-y

Godwin, A. (2016). The Development of a Measure of Engineering Identity. Paper presented at ASEE 123rd Conference & Exposition, New Orleans, June 26–29.

Godwin, A., & Potvin, G. (2017). Pushing and pulling sara: A case study of the contrasting influences of high school and university experiences on engineering Agency, Identity, and participation. *Journal of Research in Science Teaching*, *54*(4), 439–462. https://doi.org/10.1002/tea.21372

Godwin, A., Potvin, G., Hazari, Z., & Lock, R. (2016). Identity, critical agency, and engineering: An affective model for predicting engineering as a career choice. *Journal of Engineering Education*, *105*(2), 312–340. https://doi.org/10.1002/jee.20118

Graham, M. J., Frederick, J., Byars-Winston, A., Hunter, A., & Handelsman, J. (2014). Increasing persistence of college students in STEM. *Science*, *341*, 6153, 1455–1456. https://doi.org/10.1126/science.1240487

Guo, X., Hao, X., Ma, J., Wang, H., & Hu, W. (2022). Examining relationships between chemistry anxiety, chemistry identity, and chemistry career choice in terms of gender: A comparative study using multigroup structural equation modelling. *Chemistry Education Research and Practice*, *23*(4), 829–843. https://doi.org/10.1039/D2RP00070A

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7th ed.,). Prentice-Hall.

Hannachi, A., & Trendafilov, N. (2017). Archetypal analysis: Mining weather and climate extremes. *Journal of Climate*, *30*(17), 6927–6944. https://doi.org/10.1175/JCLI-D-16-0798.1

Hanson, S. L., Sykes, M., & Berneth Pena, L. (2018). Gender equity in science: The global context. *International Journal of Social Science Studies*, *6*(1), 33–47. https://doi.org/10.11114/ijsss.v6i1.2704

Harring, J. R., & Hodis, F. A. (2016). Mixture modeling: Applications in educational psychology. *Educational Psychologist*, *51*(3–4), 354–367. https://doi.org/10.1080/00461520.2016.1207176

Hazari, Z., Sonnert, G., Sadler, P. M., & Shanahan, M. C. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study. *Journal of Research in Science Teaching*, *47*(8), 978–1003. https://doi.org/10.1002/tea.20363

Hazari, Z., Chari, D., Potvin, G., & Brewe, E. (2020). The context dependence of physics identity: Examining the role of performance/competence, recognition, interest, and sense of belonging for lower and upper female physics undergraduates. *Journal of Research in Science Teaching*, *57*, 1583–1607.

Hill, P., Spiegel, A., & McQuillan, J., and J. Diamond (2018). Discovery Orientation, cognitive Schemas, and disparities in science identity in early adolescence. *Sociological Perspectives*, *61*(1), 99–125. https://doi.org/10.1177/0731121417724774

Howard, M. C., & Hoffman, M. E. (2018). Variable-Centered, Person-Centered, and Person-Specific approaches: Where theory Meets the method. *Organizational Research Methods*, *21*(4), 846–876. https://doi.org/10.1177/1094428117744021

Jaiswal, A., Magana, A. J., & Ward, M. D. (2022). Characterizing the identity formation and sense of belonging of the students enrolled in a data science learning community. *Education Sciences*, *12*(10). https://doi.org/10.3390/educsci12100731

Jiang, Z., Tang, X., & Tan, L. (2025). STEM identity and STEM career intention: A meta-analysis. *International Journal of STEM Education*, *12*, 57. https://doi.org/10.1186/s40594-025-00578-8. Su R. & Wei B.

Jones, B. D., Ruff, C., & Paretti, M. C. (2013). The impact of engineering identification and stereotypes on undergraduate women's achievement and persistence in engineering. *Social Psychology of Education*, *16*, 471–493. https://doi.org/10.1007/s11218-013-9222-x

Kang, H., Calabrese Barton, A., Tan, E., Simpkins, S., Rhee, H., & Turner, C. (2019). How do middle school girls of color develop STEM identities? Middle school girls' participation in science activities and identification with STEM careers. *Science Education*, *103*(2), 418–439. https://doi.org/10.1002/sce.21492

Kessels, U. (2005). Fitting into the stereotype: How gender-stereotyped perceptions of prototypic peers relate to liking for school subjects. *European Journal of Psychology of Education*, *20*(3), 309–323.

Kim, A. Y., Sinatra, G. M., & Seyranian, V. (2018). Developing a STEM identity among young women: A social identity perspective. *Review of Educational Research*, *88*(4), 589–625. https://doi.org/10.3102/0034654318779957

Kocsis, Á., & Molnár, G. (2025). Factors influencing academic performance and dropout rates in higher education. *Oxford Review of Education*, *51*, 3, 414–432. https://doi.org/10.1080/03054985.2024.2316616

Kozoll, R. H., & Osborne, M. D. (2004). Finding meaning in science: Lifeworld, Identity, and self. *Science Education*, *88*(2), 157–181. https://doi.org/10.1002/sce.10108

Kuchynka, S., Findley-Van Nostrand, D., & Pollenz, R. S. (2019). Evaluating psychosocial mechanisms underlying STEM persistence in undergraduates: Scalability and longitudinal analysis of three cohorts from a six-day pre-college engagement STEM academy program. *CBE—Life Sciences Education*, *18*(3). https://doi.org/10.1187/cbe.19-01-0028

Laursen, B., & Hoff, E. (2006). Person-centered and Variable-Centered approaches to longitudinal data. *Merrill-Palmer Quarterly*, *52*(3), 377–389. https://doi.org/10.1353/mpq.2006.002

Lee, Y., & Yoo, S. (2020). Individual profiles and team classes of the climate for creativity: A multilevel latent profile analysis. *Creativity and Innovation Management*, *29*()3), 438–452. https://doi.org/10.1111/caim.12371

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive career theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, *45*, 79–122.

Li, K. C., & Wong, B. T. M. (2019). Factors related to student persistence in open universities: Changes over the years. *International Review of Research in Open & Distributed Learning*, *20*(4), 132–151. https://doi.org/10.19173/irrodl.v20i4.4103

Li, N., Fan, W., Wiesner, M., Arbona, C., & Hein, S. (2021). Adapting the Utrecht management of identity commitments scale to assess engineering identity formation. *Journal of Engineering Education*, *110*(4), 885–901. https://doi.org/10.1002/jee.20417

Lin, J. C. P., Moon, S., Guan, S. S. A., Kwan, P. P., Flores, G. E., & Chavira, G. (2025). Factors influencing science career intention: The power of counterspace. *Innovative Higher Education*, *50*, 641–663. https://doi.org/10.1007/s10755-024-09752-2

Liong, M., Yeung, D. Y., Cheng, G. H. L., & Cheung, R. Y. H. (2023). Profiles of ICT identity and their associations with female high school students' intention to study and work in ICT: A mixed-methods approach. *Computers and Education*, 195. https://doi.org/10.1016/j.compedu.2022.104722

Litalien, D., Gillet, N., Gagné, M., Ratelle, C. F., & Morin, A. J. S. (2019). Self-determined motivation profiles among undergraduate students: A robust test of profile similarity as a function of gender and age. *Learning and Individual Differences*, *70*, 39–52. https://doi.org/10.1016/j.lindif.2019.01.005

Liu, S., Xu, S., Li, Q., Xiao, H., & Zhou, S. (2023). Development and validation of an instrument to assess students' science, technology, engineering, and mathematics identity. *Physical Review Physics Education Research*, *19*, 010138.

Lockhart, M. E., & Rambo-Hernandez, K. (2024). Investigating engineering identity development and stability amongst first-year engineering students: A person-centred approach. *European Journal of Engineering Education*, *49*(3), 411–433. https://doi.org/10.1080/03043797.2023.2262412

Lockhart, M. E., Kwok, O. M., Yoon, M., & Wong, R. (2022). An important component to investigating STEM persistence: The development and validation of the science identity (SciID) scale. *International Journal of STEM Education*, *9*, 34.

London, B., Rosenthal, L., Levy, S., & Lobel, M. (2011). The influences of perceived identity compatibility and social support on women in nontraditional fields during the college transition. *Basic and Applied Social Psychology*, *33*, 304–321. https://doi.org/10.1080/01973533.2011.614166

Magnusson, D. (2003). The person approach: Concepts, measurement Models, and research strategy. *New Directions for Child and Adolescent Development*, *101*, 3–23. https://doi.org/10.1002/cd.79

Mahadeo, J., Hazari, Z., & Potvin, G. (2020). Developing a computing identity framework: Understanding computer science and information technology career choice. *ACM Transactions on Computing Education*, *20*(1), 1–14.

Mäkikangas, A., Tolvanen, A., Aunola, K., Feldt, T., Mauno, S., & Kinnunen, U. (2018). Multilevel latent profile analysis with covariates: Identifying job characteristics profiles in hierarchical data as an example. *Organizational Research Methods*, *21*, 931–954.

Marinetti, S., Finesso, L., & Marsilio, E. (2007). Archetypes and principal components of an IR image sequence. *Infrared Physics & Technology*, *49*(3), 272–276. https://doi.org/10.1016/j.infrared.2006.06.017

Marsh, H. W., Ludtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of selfconcept. *Structural Equation Modeling*, *16*, 191–225.

Marshman, E., Kalender, Z. Y., Nokes-Malach, T., Schunn, C., & Singh, C. (2018). Female students with a's have similar physics self-efficacy as male students with c's in introductory courses: A cause for alarm? *Physical Review Physics Education Research*, *14*(2), 020123.

Martin-Hansen, L. (2018). Examining ways to meaningfully support students in STEM. *International Journal of STEM Education*, *5*(1), 53–56. https://doi.org/10.1186/s40594-018-0150-3

Matthews, J. S., Banerjee, M., & Lauermann, F. (2014). Academic identity formation and motivation among ethnic minority adolescents: The role of the self between internal and external perceptions of identity. *Child Development*, *85*(6), 2355–2373.

Matusovich, H. M., Streveler, R. A., & Miller, R. L. (2010). Why do students choose engineering? A qualitative, longitudinal investigation of students' motivational values. *Journal of Engineering Education*, *99*(4), 289–303.

McComas, W. F., & Burgin, S. R. (2020). A critique of STEM education. *Science & Education*, *29*, 805–829. https://doi.org/10.1007/s11191-020-00138-2

Miles, J. A., & Naumann, S. E. (2021). Science self-efficacy in the relationship between gender & science identity. *International Journal of Science Education*, *43*, 17, 2769–2790. https://doi.org/10.1080/09500693.2021.1986647

Miles, S. J., & Vela, K. N. (2025). Cultivating STEM identity and dispositions: Using virtual STEM role models in high school. *Journal for STEM Education Research*, *8*, 500–529. https://doi.org/10.1007/s41979-024-00137-w

Milite, S., Caravagna, G., & Sottoriva, A. (2025). MIDAA: Deep archetypal analysis for interpretable multi-omic data integration based on biological principles. *Genome Biology*, *26*, 90. https://doi.org/10.1186/s13059-025-03530-9

Morin, A. J., Morizot, J., Boudrias, J. S., & Madore, I. (2011). A multi-foci person-centered perspective on workplace affective commitment: A latent profile/factor mixture analysis. *Organizational Research Methods*, *14*, 58–90.

Morin, A. J., Bujacz, A., & Gagné, M. (2018). Person-centered methodologies in the organizational sciences: Introduction to the feature topic. *Organizational Research Methods*, *21*, 803–813.

Munoz, A. Z., Viveros, M., Barajas-Salazar, B., Tokman, M., Oka, L., Thompson, K., Rutter, E. M., Tran, K., Kim, C., de Sousa, C., Lei, Y., Almeida, M., & Menon, R. (2025). Roles of mathematics-related psychological factors in STEM sense of belonging and identity: A structural equation modeling analysis. *International Journal of STEM Education*, *Ed 12*, 68. https://doi.org/10.1186/s40594-025-00586-8

Ovid, D., Abrams, L., Carlson, T., Dieter, M., Flores, P., Frischer, D., Goolish, J., Bernt, M. L. F., Lancaster, A., Lipski, C., Luna, J. V., Luong, L. M. C., Mullin, M., Newman, M. J., Quintero, C., Reis, J., Robinson, F., Ross, A. J., Simon, H., Souza, G., Taylor, J., Ward, K. E., White, Y. L., Witkop, E., Yang, C., Zenilman, A., Zhang, E., Schinske, J. N., & Tanner, K. D. (2023). Scientist spotlights in secondary schools: Student shifts in multiple measures related to science identity after receiving written assignments. *CBE—Life Sciences Education*, *22*(2). https://doi.org/10.1187/cbe.22-07-0149

Palazzo, L., Iannario, M., & Palumbo, F. (2024). Integrated assessment of financial knowledge through a latent profile analysis. *Behaviormetrika*, *51*, 319–339. https://doi.org/10.1007/s41237-023-00217-y

Passeggia, R., Testa, I., Esposito, G., De Luca Picione, R., Ragozini, G., & Freda, M. F. (2023). Examining the relation between First-year university students' intention to Drop-out and academic engagement: The role of Motivation, subjective Well-being and retrospective judgements of school experience. *Innovative Higher Education*, *48*, 837–859. https://doi.org/10.1007/s10755-023-09674-5

Passeggia, R., Testa, I., Palazzo, L., & Esposito, G. (2025). Challenges of the transition to university: Identifying psychological and academic profiles of STEM university students. *Journal of Applied Research in Higher Education*. https://doi.org/10.1108/JARHE-05-2025-0365

Patrick, A. D., Borrego, M., & Prybutok, A. N. (2018). Predicting persistence in engineering through an engineering identity scale. *International Journal of Engineering Education*, *34*(2), 351–363.

Pfeuffer, J., Stoeger, H., & Ziegler, A. (2025). Amplified social inheritance in STEM: Parental careers shape adolescent identity and choice intentions. *International Journal of Adolescence and Youth*, *30*(1), 2551107. https://doi.org/10.1080/02673843.2025.2551107

Phan, H. P., Ngu, B. H., & Alrashidi, O. (2016). Role of student well-being: A study using structural equation modeling. *Psychological Reports*, *119*(1), 77–105. https://doi.org/10.1177/0033294116656819

Ragozini, G., Palumbo, F., & D'Esposito, M. R. (2017). Archetypal analysis for data-driven prototype identification. *Statistical Analysis and Data Mining: the ASA Data Science Journal*, *10*(1), 6–20. https://doi.org/10.1002/sam.11325

Reinholz, D. L., Matz, R. L., Cole, R., & Apkarian, N. (2019). STEM is not a monolith: A preliminary analysis of variations in STEM disciplinary cultures and implications for change. *CBE—Life Sciences Education*, *18*(4), ar50. https://doi.org/10.1187/cbe.19-02-0038

Renninger, K. A., & Hidi, S. E. (2021). Interest development, self-related information processing, and practice. *Theory into Practice*, *61*(1), 23–34. https://doi.org/10.1080/00405841.2021.1932159

Richter, A., & Kjellgren, B. (2024). Engineers of the future: Student perspectives on integrating global competence in their education. *European Journal of Engineering Education*, *49*(3), 474–491. https://doi.org/10.1080/03043797.2023.2298319

Robinson, K. A., Perez, T., Nuttall, A. K., Roseth, C. J., & Linnenbrink-Garcia, L. (2018). From science student to scientist: Predictors and outcomes of heterogeneous science identity trajectories in college. *Developmental Psychology*, *54*(10), 1977–1992. https://doi.org/10.1037/dev0000567

Robinson, K. A., Perez, T., Carmel, J. H., & Linnenbrink-Garcia, L. (2019). Science identity development trajectories in a gateway college chemistry course: Predictors and relations to achievement and STEM pursuit. *Contemporary Educational Psychology*, *56*, 180–192.

Rocca, A., & Quintano, C. (2024). Success stems from STEM fields: An analysis of Italian graduates. *Italian Economic Journal*, *10*, 761–793. https://doi.org/10.1007/s40797-023-00255-1

Seth, S., & Eugster, M. J. A. (2016). Probabilistic archetypal analysis. *Machine Learning*, *102*, 85–113.

Seyranian, V., Madva, A., Duong, N., Abramzon, N., Tibbetts, Y., & Harackiewicz, J. M. (2018). The longitudinal effects of STEM identity and gender on flourishing and achievement in college physics. *International Journal of STEM Education*, *5*(1), 40–14. https://doi.org/10.1186/s40594-018-0137-0

Shernoff, D. J., Kelly, S., Tonks, S. M., Anderson, B., Cavanagh, R. F., Sinha, S., & Abdi, B. (2016). Student engagement as a function of environmental complexity in high school classrooms. *Learning and Instruction*, *43*, 52–60. https://doi.org/10.1016/j.learninstruc.2015.12.003

Silva, S. D., Thompson, R., & Atkinson, M. B. (2025). BioCheM-ID: Measuring undergraduate STEM identity across biology, chemistry, and mathematics and Understanding the impacts of informal STEM experiences. *Research in Science & Technological Education*. https://doi.org/10.1080/02635143.2025.2462908

Simpson, A., McCann, J., Miroff, L. E., & McGee, T. (2025). The complex STEM identity development of middle school girls in a culturally infused afterschool program. *Journal for STEM Education Research*. https://doi.org/10.1007/s41979-025-00162-3

Singer, A., Montgomery, G., & Schmoll, S. (2020). How to foster the formation of STEM identity: Studying diversity in an authentic learning environment. *International Journal of STEM Education*, *7*(1), 1–12. https://doi.org/10.1186/s40594-020-00254-z

Smith, J. L., Brown, E. R., Thoman, D. B., & Deemer, E. D. (2015). Losing its expected communal value: How stereotype threat undermines women's identity as research scientists. *Social Psychology of Education*, *18*(3), 443–466. https://doi.org/10.1007/s11218-015-9296-8

Sonnert, G., Reid, T., Sunbury, S., & Sadler, P. (2025). How do formal and informal science learning experiences during high school shape students' career interest and STEM identity? *International Journal of STEM Education*, *12*, 55. https://doi.org/10.1186/s40594-025-00568-w

Starr, C. R. (2018). I'm not a science nerd! STEM stereotypes, identity, and motivation among undergraduate women. *Psychology of Women Quarterly*, *42*(4), 489–503. https://doi.org/10.1177/0361684318793848

Starr, C. R., Hunter, L., Dunkin, R., Honig, S., Palomino, R., & Leaper, C. (2020). Engaging in science practices in classrooms predicts increases in undergraduates' STEM motivation, identity, and achievement: A short-term longitudinal study. *Journal of Research in Science Teaching*, *57*(7), 1093–1118. https://doi.org/10.1002/tea.21623

Stringer, K., Mace, K., Clark, T., & Donahue, T. (2020). STEM focused extracurricular programs: Who's in them and do they change STEM identity and motivation? *Research in Science & Technological Education*, *38*(4), 507–522. https://doi.org/10.1080/02635143.2019.1662388

Sundstrom, M., & Holmes, N. G. (2025). Bias in physics peer recognition does not explain gaps in perceived peer recognition. *Nature Physics*, *21*, 524–529. https://doi.org/10.1038/s41567-025-02789-w

Tajfel, H. (1981). *Human groups and social categories*. Cambridge University Press.

Thompson, R. J. Jr., Schmid, L., Mburi, M., Dowd, J. E., Finkenstaedt-Quinn, S. A., Shultz, G. V., Gere, R., Schiff, A., Flash, L. A. P., & Reynolds, J. A. (2023). Diversity of undergraduates in STEM courses: Individual and demographic differences in changes in self-efficacy, epistemic beliefs, and intrapersonal attribute profiles. *Studies in Higher Education*. https://doi.org/10.1080/03075079.2023.2250385

Tripp, J., & Liu, X. (2024). Towards defining STEM professional identity: A qualitative survey study. *Journal for STEM Education Research*. https://doi.org/10.1007/s41979-024-00131-2

Verdín, D. (2021). The power of interest: Minoritized women's interest in engineering fosters persistence beliefs beyond belongingness and engineering identity. *International Journal of STEM Education*, *8*, 33. https://doi.org/10.1186/s40594-021-00292-1

Vincent-Ruz, P., & Schunn, C. D. (2018). The nature of science identity and its role as the driver of student choices. *International Journal of STEM Education*, *5*(1), 48–12. https://doi.org/10.1186/s40594-018-0140-5

Williams, M., & George-Jackson, C. E. (2014). Using and doing science: Gender, self-efficacy, and science identity of undergraduate students in STEM. *Journal of Women and Minorities in Science and Engineering*, *20*(2), 99–126. https://doi.org/10.1615/JWomenMinorScienEng.2014004477

Zhoc, K. C. H., Webster, B. J., King, R. B., Li, J. C. H., & Chung, T. S. H. (2019). Higher education student engagement scale (HESES): Development and psychometric evidence. *Research in Higher Education*, *60*, 219–244. https://doi.org/10.1007/s11162-018-9510-6

Publisher's note