USING DATA INSIDE AND OUTSIDE THE CLASSROOMS

IX Seminar "Data from and for educational system: tools for research and teaching"

edited by
Patrizia Falzetti

FrancoAngeli

INVALSI

INVALSI PER LA RICERCA
STUDI E RICERCHE

INVALSI PER LA RICERCA

La collana Open Access INVALSI PER LA RICERCA si pone come obiettivo la diffusione degli esiti delle attività di ricerca promosse dall'Istituto, favorendo lo scambio di esperienze e conoscenze con il mondo accademico e scolastico.
La collana è articolata in tre sezioni: "Studi e ricerche", i cui contributi sono sottoposti a revisione in doppio cieco, "Percorsi e strumenti", di taglio più divulgativo o di approfondimento, sottoposta a singolo referaggio, e "Rapporti di ricerca e sperimentazioni", le cui pubblicazioni riguardano le attività di ricerca e sperimentazione dell'Istituto e non sono sottoposte a revisione.

OPEN ACCESS
la soluzione FrancoAngeli

Il presente volume è pubblicato in open access, ossia il file dell'intero lavoro è liberamente scaricabile dalla piattaforma **FrancoAngeli Open Access** (http://bit.ly/francoangeli-oa).

FrancoAngeli Open Access è la piattaforma per pubblicare articoli e monografie, rispettando gli standard etici e qualitativi e la messa a disposizione dei contenuti ad accesso aperto. Oltre a garantire il deposito nei maggiori archivi e repository internazionali OA, la sua integrazione con tutto il ricco catalogo di riviste e collane FrancoAngeli massimizza la visibilità, favorisce facilità di ricerca per l'utente e possibilità di impatto per l'autore.

Per saperne di più: Pubblica con noi

 ISBN 9788835191773

USING DATA INSIDE AND OUTSIDE THE CLASSROOMS

IX Seminar "Data from and for educational system: tools for research and teaching"

edited by
Patrizia Falzetti

INVALSI

FrancoAngeli

The opinions expressed are solely of the author(s). In no case should they be considered or construed as representing an official position of INVALSI.

Assistant Editor: Francesca Leggi.

Isbn: 9788835185604

Isbn e-book Open Access: 9788835191773

Index

Introduction

by Patrizia Falzetti

In an increasingly complex educational landscape, access to and conscious use of data is key to understanding learning processes, reducing inequalities and supporting informed decision-making at instructional, institutional and policy levels. This volume, which collects a selection of contributions presented at the IX Seminar, "Data from and for Educational System: Tools for Research and Teaching" (Rome, 17-19 October 2024), explores how different data sources can contribute to a deeper understanding of how educational systems function, helping to improve students' educational and professional opportunities.

Although the four chapters differ in terms of their methods, research questions and reference contexts, they are united by a common perspective: the importance of data as both an interpretative and an operational tool.

The first chapter analyses the responses of Year 5 primary school students to the INVALSI English reading test at the basic (A1) level. Using the Rasch model enables us to observe how the difficulty of these items is stably distributed along the latent trait, providing valuable insights into the cognitive processes engaged by young readers and the cognitive demands of global comprehension.

The second chapter uses data from PISA 2022 to explore the interactions among educational expectations, socioeconomic status, language of instruction, and reading literacy among fifteen-year-old students in the Republic of North Macedonia.

The third chapter introduces the UNI.CO model, an innovative system which integrates administrative data from universities with employment data in order to trace the longitudinal transitions of graduates into the labour market. The results highlight structural challenges in the Italian labour market and offer a concrete example of how integrated information systems can support university planning and active labour market policies.

The fourth and final chapter focuses on the potential of machine learning to predict university dropout using a comprehensive national dataset combining school, university and INVALSI assessment data. The analysis demonstrates how these models can identify students at risk of dropping out early on, using a set of predictors that includes demographic, educational, and performance variables. The aim is to demonstrate the effectiveness of supervised learning techniques and their potential impact on educational policies and guidance interventions.

Together, the four chapters offer a multifaceted perspective on the value of data in understanding educational phenomena, both in and out of the classroom. They demonstrate how various methodological approaches, such as psychometrics, multivariate statistical analysis, administrative data integration and predictive modelling, can converge towards the shared objective of developing more equitable, informed and responsive educational systems that better address the needs of students and society.

1. Using INVALSI data to explore beginner L2 English learners' performance while reading to grasp the general meaning of a text

by Clelia Cascella, Francesca La Russa*

Second language (L2) reading comprehension is a multifaceted cognitive process involving various dimensions. This study adopts Khalifa and Weir's (2009) reading model, which differentiates between lower-level processes – from word recognition to establishing propositional meaning – and higher-level processes – from inferencing to creating (inter)textual representations. While proficient readers typically automate lower-level processes, beginner L2 learners may not have fully automated them due to their limited L2 proficiency. So, they spend more cognitive resources on lower-level processes, leaving fewer resources for higher-level operations.

The study investigates the responses of grade 5 students to INVALSI English reading tests at the A1 level in the 2022/2023 school year, especially focusing on items that require students to grasp the general meaning of short texts

* The present work stems from a close cooperation between the two authors. The authors' names are listed in alphabetical order. For the specific concerns of the Italian Academy, Clelia Cascella is responsible for paragraphs § 4, § 5, and § 6; Francesca La Russa is responsible for paragraphs § 1, § 2 and § 3. Paragraph § 7 has been conjointly written by the authors.

(gist). Reading for gist involves reconstructing the overall meaning of a text based on key words or sentences. Given students' beginner proficiency, the hypothesis is that gist reading poses greater challenges compared to other types of reading tasks.

INVALSI data were analyzed by using the Rasch model. Items' location along the latent trait showed that gist items consistently exhibit medium difficulty and form a homogeneous block along the latent trait. This suggests that the cognitive load required for gist reading remains stable and is intrinsically tied to the reading strategy.

La comprensione della lettura in lingua seconda (L2) è un processo cognitivo complesso e multidimensionale. Il modello di lettura proposto da Khalifa e Weir (2009) distingue tra processi cognitivi di livello inferiore – dal riconoscimento della parola alla definizione del significato proposizionale – e processi di livello superiore, dall'inferenza alla costruzione di rappresentazioni (inter)testuali. Mentre i lettori competenti automatizzano i processi di livello inferiore, gli apprendenti principianti di una L2, a causa della limitata competenza linguistica, dedicano a questi processi una quota significativa delle loro risorse cognitive, lasciando meno risorse per le operazioni più complesse.

Lo studio analizza le risposte degli studenti di quinta primaria ai test INVALSI di Inglese reading di livello A1, somministrati nell'anno scolastico 2022/2023. Particolare attenzione è dedicata agli item che richiedono di cogliere il significato globale (gist) di brevi testi, ricostruendo la macrostruttura del testo a partire da parole o frasi chiave. L'ipotesi è che, dato il livello principiante degli studenti, questi item risultino più difficili rispetto ad altri. I dati INVALSI sono stati analizzati utilizzando il modello di Rasch. Il posizionamento degli item lungo il tratto latente indica che gli *item gist presentano una difficoltà media costante e si distribuiscono in modo omogeneo lungo il tratto latente. Questo risultato suggerisce che il carico cognitivo associato a tali item sia stabile e strettamente legato alla strategia di lettura.*

1. Introduction

The English tests administered by the Italian national institute for the evaluation of the educational system (INVALSI), currently administered on a census basis to grade 5, 8 and 13 Italian students, have been introduced by Legislative Decree no. 62 of 2017. The Decree requires INVALSI to administer achievement tests on receptive skills (reading and listening) consistent with the *Common European Framework of Reference for Languages* – CEFR (Council of Europe, 2001, 2020).

The CEFR, developed by the Council of Europe in 2001 and then updated by a Complementary Volume published in 2020, is a reference tool that provides a comprehensive descriptive scheme «of language competence and a set of common reference levels (from A1 to C2), defined in scales of descriptors, and more options for the design of curricula and programs for the promotion of plurilingual and intercultural education» (Council of Europe, 2020, p. 25). This descriptive scheme can be used as a basis for the transparent definition of curriculum objectives, standards and criteria for assessment starting from what users/learners need to do with the language.

According to a socio-cognitive framework (Weir, 2005), language use involves both social factors (it is a social phenomenon rather than a purely linguistic one) and cognitive factors. However, the CEFR remains mainly a descriptive scheme of linguistic competence that only indirectly provides some indications on the cognitive processes that language users/learners can face at each proficiency level in various linguistic-communicative activities. For reading tests, this scheme should thus be integrated with a description of the cognitive processes that support written reception.

This study adopts Khalifa and Weir's (2009) model of reading, which differentiates between lower-level cognitive processes, leading to reconstructing meaning at the local sentence level, and higher-level processes, leading to the reconstruction of a text's global meaning. Specifically, this study examines the responses of grade 5 students to INVALSI English reading tests at the A1 level, with a particular focus on items designed to assess the ability to read for gist, i.e. understanding the overall meaning of short texts. Given the beginner proficiency level of the students, the hypothesis posits that reading for gist may present greater difficulty compared to other types of reading tasks.

The following sections outline the cognitive model of reading that underpins this study with a focus on reading for gist (§ 2), the research aims and hypothesis (§ 3), the analytical methods employed (§ 4), a detailed presentation of the results (§ 5) and their discussion (§ 6), conclusions, limitations, and guidelines for future research (§ 7).

2. Background

Reading is a complex cognitive process involving multiple factors and dimensions. This complexity becomes even more pronounced in second language (L2) reading, where linguistic, cognitive, and cultural variables – such as L2 exposure, overall L2 proficiency, and processing skills – play a pivotal role in comprehension. The interplay of all these factors makes it difficult to establish a unified conceptualization of such a multifaceted phenomenon (Grabe and Yamashita, 2009).

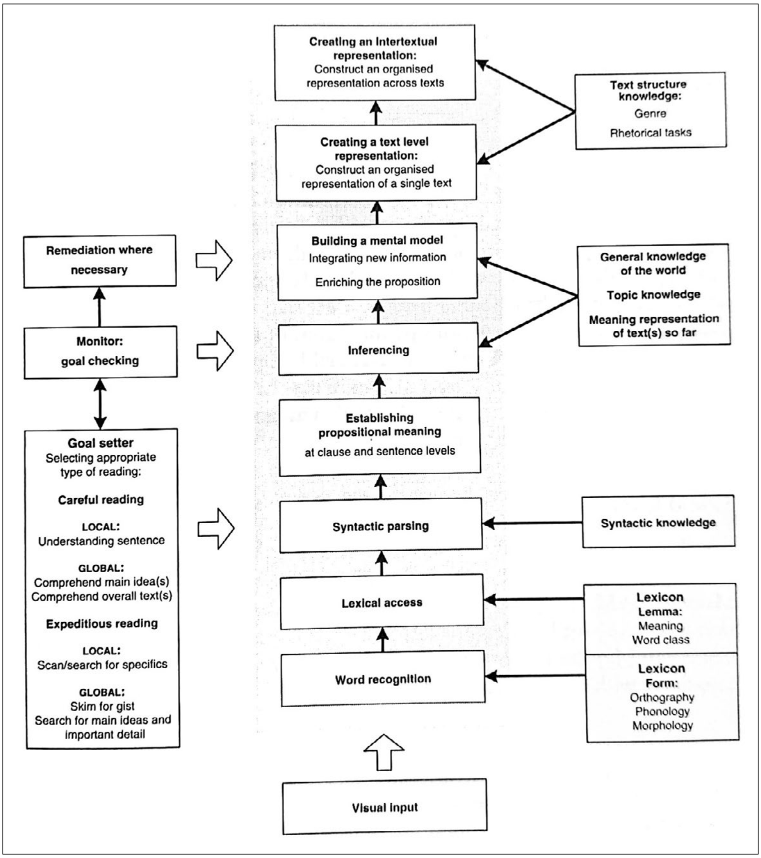

Fig. 1 — The model of reading by Khalifa and Weir (2009, p. 43)

Nevertheless, a balance must be struck between capturing the dynamic nature of reading and providing a framework that can inform assessment practices. This study adopts as a reference the reading model proposed by Khalifa and Weir (2009) shown in Figure 1. Unlike other models that are not explicitly tailored to address L2 reading processes or their application in assessment contexts, Khalifa and Weir's model builds on insights from cognitive psychology related to first language (L1) reading and adapts them specifically for L2 learners. Furthermore, it places particular emphasis on cognitive processing as a theoretical foundation for the cognitive validity of reading tests.

As illustrated in Figure 1, the model accounts for the cognitive and metacognitive processes as well as the knowledge that underpin reading.

The reading process begins with the metacognitive activity of goal setting. In this phase, the reader selects the purpose and reading method to be implemented. A distinction is made between careful reading, a slow, linear and incremental reading aimed at understanding the complete meaning of a sentence (local level) or of an entire text (global level); expeditious reading, i.e. a rapid and selective reading for finding the desired information within the text. At a local level, this type of reading can be used to capture specific information (scanning). At a global level, it is used to quickly identify the structure of the text, the main ideas or the overall meaning (skimming). Other metacognitive activities (i.e. monitoring and remediating) might be implemented throughout the process to check understanding in the various stages of reading, remediating where necessary.

As for the cognitive processes, once the visual input, i.e. the text, has been received, the first process is that of word recognition which involves the identification of the orthographic, phonological and morphological form of the word. Lexical access follows. Field (2004, p. 151) defines lexical access as «retrieval of a lexical entry from the lexicon, containing stored information about a word's form and its meaning». Information on the part of speech to which the word belongs, on the syntactic structures in which it can appear and on the range of its possible senses are thus retrieved. The next step is syntactic parsing, in which words are organized into phrases and larger units. It is then possible to establish a literal interpretation of the propositional meaning at clause or sentence level.

However, literal meaning may not be sufficient to understand the complete meaning of a text and the reader must add external knowledge (general knowledge, topic knowledge, representation of the meaning of the text) to understand it at full. It is therefore necessary to implement other cognitive processes of a higher cognitive level. The first of these is inferencing, a creative process by which information not explicitly expressed is added. Inference can happen at different levels depending on whether it is a matter of identifying the antecedent of a pronoun, inferring the meaning of an unknown word, recognizing the implications of a sentence, reconstructing the information left implicit by the author, reconstructing the connections between ideas to impose coherence, and so on. The next phase consists in building a mental model of the text, integrating the new information with the previous one. Subsequently, thanks to the knowledge of textual genres and rhetorical conventions, a text-level representation is created. This involves understanding the relationships between information and recognizing their hierarchical structure, distinguishing between main and secondary ideas. The same happens at an intertextual level if more than one text is involved, in this case however it is a matter of grasping the relationships between different texts.

A distinction can be drawn between lower-level cognitive processes – from word recognition up to establishing propositional meaning – which require the reader to focus on the linguistic code to decode the text and reconstruct its literal meaning, and higher-level cognitive processes –

from inferencing up to creating an (inter)textual representation – that enable a comprehensive understanding of the text's macrostructure and global meaning.

The cognitive processes employed while reading can be influenced by the reader's L2 proficiency level. In this regard, the theoretical framework of the PISA 2025 *Foreign Language Assessment* survey (OECD, 2021, p. 44) states: «From Levels Pre-A1 to A2, the focus on short, simple texts of a concrete nature suggests that these learners are primarily able to cope with the lower level reading processes (those up to establishing propositional meaning), with the main purpose employed being careful reading at the local level due to their limited grammatical and lexical knowledge».

Lower-level cognitive processes are more dependent on knowledge of the linguistic code (Cohen and Upton, 2006). Orthographic, phonological and morphological knowledge of the L2 is needed for word recognition; lexical knowledge of the L2 is required for lexical access, while knowledge of L2 grammar and syntax is needed for syntactic parsing. These processes are usually automatically performed by proficient readers. However, beginner L2 learners, due to their limited linguistic proficiency, may struggle to automate them. As a result, they may allocate a significant portion of their attentional resources to completing lower-level processes, leaving fewer resources available for higher-level ones (Nassaji, 2014).

2.1. Reading for gist

Reading for gist is a skill that typically involves reconstructing the text's macrostructure based on some details (keywords or sentences), thus identifying the macro-propositions that encapsulate the text's overall meaning.

In this paragraph, we present some examples of items testing reading for gist (Fig. 2), taken from the INVALSI achievement test administered in the school year 2022/2023. These items require test takers to read some short texts about different shops and match them with the corresponding questions "Where can you go if you want to...?".

- **A** **Bake and Bite**
  Delicious sandwiches, cakes and biscuits with hot tea or coffee.
- **B** **Animal Kingdom**
  Top quality food for dogs and cats, and lots of toys for your animal friends.
- **C** **Paperchase**
  We have school bags, diaries and a big selection of pencils and pens. Open from 9:00am to 7:00pm.
- **D** **Garden Home**
  Come inside and see our fresh flowers for every occasion and every day.
- **E** **The Gaming Store**
  Everything from video games to model airplanes, card games to puzzles. Something for everybody!
- **F** **The Spirit of India**
  Come and eat real Indian food! Sit outside and enjoy perfect rice and curry.
- **G** **Kids and Co.**
  Fantastic clothes for children from 3 to 12 years old with a new summer selection for the beach.

Where can you go if you want to

0	buy a T-shirt for a child?	
Q1	find many things to play with?	
Q2	have a good breakfast?	
Q3	buy everything for your lessons?	
Q4	eat food from another country?	

Fig. 2 – A1 gist items from task 4 grade 5 INVALSI English reading test 2022/2023

In Question 1 (Q1), the target information is found in short text E, which consists of two sentences. A superficial association between *play* in the question and the title *Gaming Store* may not be sufficient for a correct answer, as text B presents a potential distractor by mentioning *lots of toys*, though for pets. Only by reading the full text E students can confirm that the store sells a variety of play-related items, including video games, puzzles, and toy airplanes.

In Question 2 (Q2), the relevant information is in short text A. Test takers might initially link *breakfast* in the question with *Bake and Bite* in the title. This association is confirmed upon reading the full text, which lists cakes, biscuits, hot tea, and coffee – typical breakfast food. Text F presents a distractor, as it also mentions food but refers to Indian cuisine.

In Question 3 (Q3), the target information is in short text C, consisting of two sentences. To answer correctly, test takers must associate *lessons* with the school-related objects mentioned in the text, such as school bags, diaries, pencils, and pens.

Finally, in Question 4 (Q4), the target information is in short text F, which also consists of two sentences. Simply recognizing *food* in both the question and the text is insufficient, as two distractors are present: text A, which refers to pet food, and text B, which is about breakfast food. To arrive at the correct answer, test takers must carefully read the full text F.

All four questions require test takers to go beyond simple word matching and engage in understanding the overall meaning of the short texts. Potential distractors demand that students carefully read the full text to confirm their initial associations, ensuring comprehension of the broader context rather than relying on isolated words to make surface-level connections.

3. Aims and research hypothesis

This study examines the responses of grade 5 students to the INVALSI English reading test. According to the national curriculum, students are expected to achieve the A1 proficiency level. As for reading comprehension, they should be able to read and comprehend short simple texts, preferably accompanied by visual aids, grasping their overall meaning and identifying familiar words and phrases (*Indicazioni nazionali*, 2012).

Particular attention is given here to the students' ability to read for gist (§ 2.1). The objective is to determine whether, given the students' limited L2 proficiency, this type of reading proves to be more challenging compared to other reading tasks aimed at testing other reading goals (§ 2).

4. Methods

4.1. Data

Every year, INVALSI administers achievement tests in English as a foreign language at both the census and sample levels. To test our research hypothesis, we analyzed data from the INVALSI sample because they are free from cheating (Longobardi, Falzetti and Pagliuca, 2018). Indeed, the data from the INVALSI sample classes are collected under the supervision of an external observer, who is responsible for verifying that all administrative procedures are implemented fairly and that there is no cheating by students (or teachers).

The results derived from these data are therefore not subject to the potential distorting effects of cheating. The results presented in the current study are based on data collected within the most recent administrative wave, that of the 2022/2023 school year.

The INVALSI English achievement test that we analyzed, like all the others administered at grade 5, consisted of five reading tasks with a range of difficulty from Pre-A1 to A1 CEFR. As in the case of INVALSI, an analysis of all the items administered was carried out using the Rasch model (1960/1980). However, given that our interest is in the A1 proficiency level and that only one of the five items administered is at the Pre-A1 level (while all the others are at the A1 level), the qualitative analysis and interpretation of the results of the Rasch analysis focused only on those items that are both gist and A1.

4.2. The analytical strategy

In the present study, we deployed the Rasch model (1960/1980) to analyze data. The use of the Rasch model is particularly appropriate for the purpose of analysing the INVALSI data. Indeed, students' reading ability is conceptualized as a unidimensional construct and operationalised through the selection of cumulative, locally independent items. Empirical verification of these characteristics is provided by deploying the Rasch model that assumes (i) unidimensionality (after which response probability does not depend on variables other than the investigated ability, θ) and (ii) local independence (after which there is no correlation between any two items), and (iii) monotonicity (according to which response probability increases with higher values of θ) (Hambleton and Swaminathan, 1985).

The idea behind the Rasch model is that the probability of answering an item correctly is a function of the student's relative ability[1], i.e. student intrinsic language ability relative to the difficulty of each item: the greater the student's ability relative to the difficulty of the item, the greater his or her probability of giving a correct answer.

In accordance with this logic, both students and items can be ordered along the same latent trait depending on their estimated ability/difficulty respectively: more able students (and more difficult items) are scaled at the top of the latent trait, while the less able students (and the less difficult items) are at the bottom of the latent trait.

The Rasch model is one of the statistical model most widely used in educational research, especially in large-scale surveys because, when the model's assumptions are met, Rasch estimates guarantee measurement invariance (e.g., Engelhard, 2009; Engelhard and Wang, 2024) and thus the comparability of results between groups of students, between groups of students with groups of items, and between groups of items. In the current study, we thus used this characteristic of the Rasch model to compare gist items with those aimed at investigating other reading strategies/goals.

Subsequent to the empirical demonstration of the goodness of data-model fit, an exploration was thus conducted of the items' hierarchy, i.e. their position along the latent trait depending on their estimated difficulty.

Such an analysis was undertaken in order to control the hypothesis that gist reading poses greater challenges compared to other types of reading tasks, thus resulting in it being more difficult than the other items, especially for readers with developing reading skills, such as A1 students at the conclusion of primary education. Consequently, assuming

[1] Ability in this context is a technical term used specifically in the psychometric field in connection with Rasch modelling. The term is used to refer to the amount of the trait (in this case, the student's) being measured.

that there are no other factors – other than students' reading ability – associated with the probability of successfully encountering each item, we expect that gist items will be scaled at the top of the items' hierarchy, as their estimated difficulty parameters will be the most elevated in comparison to all the other items.

In addition, for further investigation of the results and support their qualitative interpretation, we analyzed the items characteristic curve (ICCs), a graph plotted within the framework of the Rasch analysis that expresses the probability of a correct response in relation to the respondent's intrinsic ability in the trait being measured.

5. Results

In this paragraph we first explored the psychometric functionality of gist items and then critically discussed the results interpreting them from a qualitative point of view. Data analysis was carried out by using Winstep (Linacre, 2023).

The following Table 1 reports on the goodness of data–model fit by means of the presentation of both INFIT MNSQ and OUTFIT MNSQ statistics[2] (Linacre, 2002). INFIT, oth-

[2] Mean square (MNSQ) fit statistics measure the randomness or bias in the measurement system, with 1.0 being the expected value. Values below 1.0 indicate predictability (overfit), while values above 1.0 indicate unpredictability (underfit). For example, those items that are semantically very similar to most of the other items administered may introduce redundancy into the scale, resulting in an item overfitting the model. Conversely, items that are inconsistent with the other administered items will underfit the model. Such a violation can threaten the validity of the scale for a number of reasons and is therefore a major concern.

Mean-squares are chi-square statistics divided by their degrees of freedom and are always positive. Values near 1.0 suggest minimal distortion. Outfit issues are less threatening but easier to manage than infit erwise known as *inlier*-sensitive fit, is a measure of the pattern of responses to items targeted at a person. OUTFIT, otherwise known as *outlier*-sensitive fit, is more responsive to items with difficulty levels far from the person.

In addition, we computed both person separation (1.46) and reliability (0.68) as well as item separation (35.76) and reliability (1.00), that provide information about the characteristics and the appropriateness of both sample of students and the items administered. Persons' reliability and separation serve distinct functions. Low person separation (< 2.0) and reliability (< 0.8) with a relevant sample suggest the instrument may not distinguish well between high and low performers, indicating a need for more items. Item reliability and separation confirm the item hierarchy. Low item separation (< 3.0) and reliability (< 0.9) imply that the sample size is inadequate for validating the item difficulty hierarchy, which in turn affects construct validity. Reliability (separation index) is indicative of the reproducibility of relative measure locations, rather than data quality. High reliability is indicative of a high probability that high measures correspond to higher actual values. Achieving high reliability necessitates a sufficiently large sample and/or low measurement error. Conversely, high person reliability necessitates a diverse sample and/or a substantial number of items, while high item reliability requires a broad item difficulty range and/or a large sample. Low item reliability frequently arises from a limited sample size. In the analysis presented in the current paper, all the computed values confirm the goodness of both the student sample and the item set.

issues. To assess the impact of misfit, suspect responses should be replaced with missing values and changes observed. All the administered items' infit statistics are close to the ideal value, with just a few exceptions that are within the tolerance interval of ± 0.20 (Engelhard, 2009; Engelhard and Wang, 2024).

Tab. 1 – Item statistics based on data collected in the 2022/2023 school year

Item	*Total score*	*Item parameter*	*Model SE*	*INFIT MNSQ*	*OUTFIT MNSQ*
TASK1D4	16,299	-2.13	0.04	0.94	1.04
TASK3D2	15,249	-0.91	0.03	1.00	1.03
TASK5D2	15,181	-0.85	0.03	0.95	0.85
TASK5D4	15,123	-0.81	0.03	1.09	1.28
TASK4D5	14,948	-0.67	0.03	0.89	0.82
TASK1D1	14,908	-0.65	0.03	1.00	1.03
TASK3D4	14,897	-0.64	0.03	1.01	0.93
TASK5D3	14,749	-0.53	0.03	0.98	0.85
TASK1D5	14,555	-0.40	0.03	0.98	0.90
TASK4D2	14,482	-0.35	0.03	0.91	0.74
TASK2D1	14,459	-0.34	0.03	0.81	0.63
TASK2D2	14,422	-0.32	0.03	0.98	0.93
TASK1D2	14,002	-0.07	0.02	0.97	0.93
TASK3D3	13,765	0.07	0.02	1.06	1.02
TASK4D1	13,736	0.08	0.02	0.98	0.98
TASK3D5	13,627	0.14	0.02	1.09	1.27
TASK4D3	13,530	0.19	0.02	0.91	0.79
TASK4D4	13,512	0.20	0.02	0.95	0.87
TASK1D3	13,425	0.25	0.02	1.06	1.10
TASK5D5	13,404	0.26	0.02	0.96	0.89
TASK2D3	13,232	0.34	0.02	0.91	0.82
TASK2D5	12,124	0.86	0.02	0.97	0.90
TASK5D6	11,691	1.04	0.02	1.15	1.24
TASK2D4	11,344	1.19	0.02	0.96	0.91
TASK5D1	10,952	1.35	0.02	1.14	1.25
TASK3D1	7,520	2.72	0.02	1.20	1.58

Notes: Person Separation: 1.46, Person Reliability: 0.68 | Item separation: 35.76, Item reliability: 1.00.
Gist items have been highlighted in bold.

The items presented in Table 1 have been ordered in accordance with the estimated difficulty of each item, from the easiest (TASK1D4) to the most difficult item (TASK3D1). This arrangement reflects the concept of the item hierarchy, which conceptualizes the placement of items along the latent trait in accordance with the extent of the property (i.e., reading ability) required to respond correctly: the lower the value, the less challenging the item, and viceversa.

A1 gist items (presented in § 2.1) show the following parameters:

– Task4_D1_GIST_MATCH: +0.08;
– Task4_D3_GIST_MATCH: +0.19;
– Task4_D4_GIST_MATCH: +0.20;
– Task4_D2_GIST_MATCH: –0.35.

Gist items are thus scaled between – 0.5 logit and 0.5 logit, where 0.00 logit[3] is the scale mean: it does not indicate a "null" value, but a medium level of difficulty[4]. Items' position along the latent trait thus shows that the gist comprehension items are all located in the same part of the la-

[3] A logit, or "log-odds unit" is an additive measurement unit used to express the difference between a person's ability and an item's difficulty on a common scale. Derived from the natural logarithm of the odds, logits form an equal interval scale, making it easier to interpret and compare measures. A logit represents the distance that increases the probability of an event by a factor of 2.718. Logits are additive and independent of specific items or people, ensuring robust and generalizable results. This allows for precise and meaningful data interpretation in Rasch analysis.

[4] In the Rasch analysis, the software Winsteps sets the scale mean at 0.00 *logits* to provide a standardized reference point for interpreting measures. This approach simplifies the comparison of person abilities and item difficulties by centering the scale around a common, neutral value. By setting the mean at 0.00 logits, Winsteps ensures that the logit scale is balanced, making it easier to understand and interpret the relative positions of persons and items.

tent trait and present a consistent level of difficulty, unlike the other items, which are distributed unevenly.

The items that students found more challenging often required specific lexical or grammatical knowledge. For instance, Fig. 3 illustrates the first item of the third task, which demonstrated a high level of difficulty (+2.72).

I have a pet animal. My pet animal is a cat. Her name is Kitty. She is big, white and brown. She has a long tail. She has four kittens. They all are very cute. All of them are white. Fish is her favourite food. I give her milk every day. She can climb trees. She likes to play on my bed. I like to play with my Kitty. I love her very much.

		True	**False**	**Not given**
0	The girl's pet is a dog		X	
Q1	Kitty is the girl's name			

Fig. 3 – TASK3_D1 from grade 5 INVALSI English reading test 2022/2023

To answer Question 1 in Figure 3, the test taker must reconstruct the anaphoric reference of the pronoun *her* and relate it to the cat mentioned in the previous sentence. The first item ("Kitty is the girl's name") is therefore incorrect. Grammatical knowledge of possessive pronouns is required to encounter this item successfully: the student must know that her can be referred to a pet, not just to a person.

Our proposed interpretation is further supported by the item characteristic curve (Fig. 4), that expresses the probability of encountering an item usefully as function of student's ability. The graph displays a continuous bold curve, representing the expected probability of encountering the item successfully for students of varying abilities, and an empirical piecewise linear curve, representing the actual answers given by students of different abilities.

It is noteworthy that the empirical curve, particularly for lower and higher ability students, exhibits a slight flattening in comparison to the theoretical curve. This phenomenon, termed under-discrimination, can be interpreted as indicating that the process of discriminating between students based on their ability (i.e. the item discrimination process) is less effective than would be predicted based on students' estimated ability, estimated on the basis of their responses to all other test items administered.

Consequently, we can hypothesise that an alternative factor – other than students' ability – such as a misconception, modifies the probability of successfully encountering this item. The hypothesis is that this "other factor" may be students' ignorance of the use of the feminine pronoun when referring to a pet rather than a person.

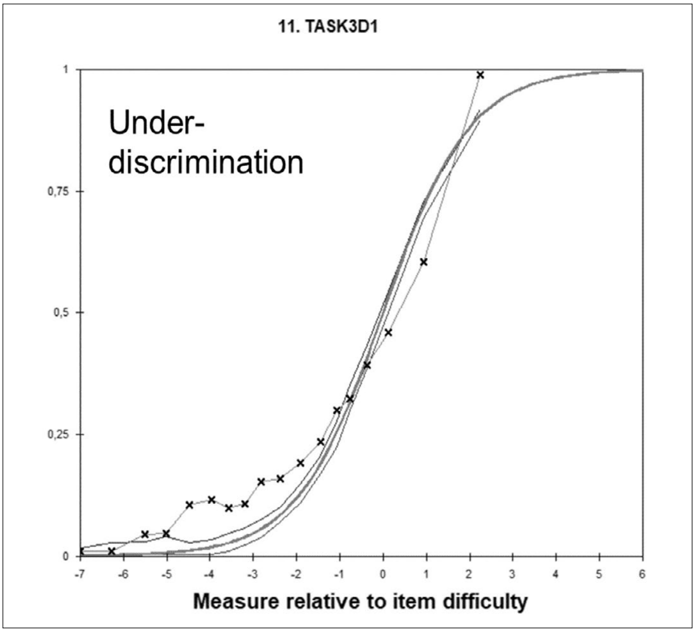

Fig. 4 – Item TASK3_D1 characteristic curve

6. Discussion

Our analysis of the INVALSI data has revealed that while items assessing gist comprehension are not necessarily more difficult than items with other reading goals, they do exhibit a consistent level of medium difficulty. Indeed, gist items are scaled along the same part of the latent trait (around 0.00 *logits*), thus forming a homogeneous block in terms of difficulty. In contrast, the results indicate that items with a different reading goal are scaled on various points of the latent trait and their positioning along the latent trait shows no regularity.

These preliminary findings thus seem to indicate that for gist items the cognitive load is pretty much always the same and might thus be intrinsically linked to the reading strategy. Conversely, for more challenging items (such as TASK3D1), the factors responsible for the difficulty do not seem to be strictly linked to the reading goal, but involve other aspects somehow associated, for example, to grammatical or lexical knowledge. This confirms that item difficulty is influenced by a broader set of factors, including lexical and syntactic complexity, the wording of the items or the characteristics of the text stimulus.

Some limitations should be acknowledged. First, data from standardized testing contexts offer precious insights into large-scale reading performance trends by providing data from a large number of students across different years, ensuring measurement consistency and generalizability, and providing strong psychometric evidence for item difficulty. However, they offer limited visibility into the cognitive processes underlying students' responses. A more fine-grained, process-oriented analysis, such as eye-tracking studies or think-aloud protocols, could provide deeper insights into how students engage with gist reading tasks and pinpoint specific areas of difficulty.

Another limitation may be the use of just one data set, because the results described in this chapter may be due to specific characteristics associated with the particular tasks administered in the booklet. This necessarily restricts the generalizability of the empirical findings, and broader conclusions would require replication or extension of the analyses to additional item sets. In this direction, subsequent studies will benefit from the data that INVALSI has collected annually at grade 5 in recent years (starting with the 2017/2018 school year). Although the national administration design has so far involved independent linear tests, it will still be possible to compare the item hierarchies obtained from different administrations to assess whether the core items consistently occupy a similar position along the scale. Importantly, the focus here is not on direct comparisons of item parameters, but on the relative position of gist items along the latent trait compared to other items targeting different reading goals. If future analyses show that gist items consistently fall within the same portion of the latent trait (as shown in our research), while other item parameters vary more widely, this would provide further evidence that gist item parameters are primarily influenced by the intended reading goal, whereas other items' parameters depend on a combination of factors, including – but not limited to – their expected reading goal. In this context, a key contribution of this chapter is the demonstration of an approach that integrates qualitative (§ 2 and § 5) and quantitative (§ 5) analyses of item functioning. While the empirical results cannot be generalized beyond the data examined here, the analytical framework is intended as a replicable example that can be applied and adapted in future research using different item sets.

7. Conclusions and guidelines for future research

The results of the data analysis contradicted our research hypothesis that gist items would be more difficult than items designed to measure other research *foci* for A1 test takers at the end of primary school in Italy.

However, such a result opens up new research questions. The results presented in this study should thus be seen as a starting point toward a more comprehensive understanding of item difficulty in reading assessments. For example, it is well known from the relevant literature that the difficulty of an item depends on a number of factors. One first promising research direction is to investigate the relationship between item and text characteristics and item difficulty using models such as the Linear Logistic Test Model (Fischer, 1973) that express item difficulty as a (linear) combination of factors. By systematically examining how lexical, syntactic, and discourse-level features contribute to the difficulty of different reading tasks, researchers can develop a more nuanced understanding of what makes certain items more challenging than others. These results might be of interest both to educational policy-makers and practitioners, as they will improve the interpretability of educational results, and to researchers and test developers, as they will make explicit the relationship between item difficulty and text/item characteristics.

Secondly, this study focused exclusively on A1-level students and their performance in reading to grasp the gist of a text. Future research could explore how readers with higher proficiency levels engage with gist reading tasks, investigating whether differences exist across proficiency levels and examining the difficulty of these tasks for more proficient readers, whose automation of lower-level processes may support more efficient gist reading.

Finally, given the increasing importance of digital reading, future research could compare gist comprehension in print versus digital forms. Differences in reading strategies, screen-based cognitive load, and navigation behaviours could influence how students process gist information in online reading contexts, a crucial consideration for modern language assessment and instruction.

In conclusion, the study provides evidence that reading for gist presents a stable cognitive challenge for beginner L2 learners, as reflected in the consistent difficulty of gist test items. These findings highlight the importance of considering reading strategies in language assessment and, given the multifaceted nature of reading, stress the importance of further research into the broader range of factors influencing reading comprehension. By further exploring the factors that influence item difficulty, future research can deepen our understanding of L2 reading development and contribute to more effective language teaching and assessment practices.

References

Cohen A.D., Upton T.A. (2006), *Strategies in responding to the new TOEFL reading tasks*, Educational Testing Services, Princeton.

Council of Europe (2001), *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge University Press, Cambridge.

Council of Europe (2020), *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors*, Council of Europe Publishing, Strasbourg.

Engelhard Jr G. (2009), "Using item response theory and model-data fit to conceptualize differential item and person functioning for students with disabilities", *Educational and Psychological Measurement*, 69 (4), pp. 585-602.

Engelhard Jr. G., Wang J. (2024), *Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences*, Routledge, New York, 2nd ed., https://doi.org/10.4324/978 1003458746.

Field J. (2004), *Psycholinguistics: The key concepts*, Psychology Press, Hove.

Fischer G. H. (1973), "The linear logistic test model as an instrument in educational research", *Acta Psychologica*, 37 (6), pp. 359-374.

Grabe W., Yamashita J. (2009), *Reading in a second language: Moving from theory to practice*, Cambridge University Press, Cambridge.

Hambleton R.K., Swaminathan H. (1985), *Item response theory: Principles and applications*, Kluwer-Nijhoff, Boston.

Italian Ministry of Education, Decree n 254 16th November 2012, *Indicazioni nazionali per il curriculo della scuola dell'infanzia e del primo ciclo di istruzione*, https://www.mim.gov.it/documents/20182/51310/DM+254_2012.pdf.

Khalifa H., Weir C.J. (2009), *Examining Reading: Research and Practice in Assessing Second Language Reading*, Studies in Language Testing 29, UCLES, Cambridge University Press, Cambridge.

Linacre J.M. (2002), "What do infit and outfit, mean-square and standardized mean", *Rasch Measurement Transactions*, 16 (2), https://www.rasch.org/rmt/rmt162f.htm.

Linacre J.M. (2023), *Winsteps® Rasch measurement computer program (Version 5.6.0)*, Winsteps.com, Portland.

Longobardi S., Falzetti P., Pagliuca M.M. (2018), "Quis custiodet ipsos custodes? How to detect and correct teacher cheating in Italian student data", *Statistical Methods & Applications*, 27, pp. 515-543.

Nassaji H. (2014), "The role and importance of lower-level processes in second language reading", *Language Teaching*, 47 (1), pp. 1-37.

OECD (2021), *PISA 2025 Foreign Language Assessment Framework*, OECD Publishing, Paris.

Rasch G. (1960/1980), *Probabilistic models for some intelligence and attainment tests (Expanded ed.)*, University of Chicago Press, Chicago.

Weir C.J. (2005), *Language testing and validation*, Palgrave Macmillan, Hampshire.

2. Career and educational expectations and reading literacy achievement: a mixed-methods study of adolescents in North Macedonia using PISA 2022

by Beti Lameva, Zaneta Chonteva

The chapter combines descriptive analyses, multivariate regression, and qualitative evidence from focus group discussions to analyse whether students' ambitious educational and career expectations, socioeconomic status, and language of instruction are aligned with their reading literacy competencies. It draws on data from the Programme for International Student Assessment (PISA) 2022 for the Republic of North Macedonia. High prevalence of ambitious educational expectations, with 71.8% of students aiming for tertiary education, and professional career expectations in fields such as information and communication technology, healthcare, and management describe students in North Macedonia. But, at the same time, we can see, a significant misalignment with academic competencies, particularly among students from lower socio-economic backgrounds or those receiving instruction in Albanian language. Regression analysis results indicate that educational and career expectations, socioeconomic background, and language of instruction significantly predict reading literacy performance. Students' narratives from the focus groups discussions further reveal that limited career guidance, undervaluation of effort, and misalignment between

expectations and skills are the main challenges in developing informed student agency. These research findings add to the literature by linking student agency, expectation formation, and measurable competencies, providing actionable insights for educational reform and career guidance in North Macedonia.

Il capitolo combina analisi descrittive, regressioni multivariate ed evidenze qualitative provenienti da discussioni in focus group per analizzare se le ambiziose aspettative educative e professionali degli studenti, lo status socioeconomico e la lingua di istruzione siano allineati con le loro competenze di lettura. Lo studio si basa sui dati del Programme for International Student Assessment (PISA) 2022 per la Repubblica della Macedonia del Nord. Un'elevata prevalenza di aspettative educative ambiziose, con il 71,8% degli studenti che mira all'istruzione terziaria, e di aspettative di carriera professionale in ambiti quali le tecnologie dell'informazione e della comunicazione, l'assistenza sanitaria e la gestione caratterizza gli studenti della Macedonia del Nord. Tuttavia, allo stesso tempo, emerge un significativo disallineamento rispetto alle competenze accademiche, in particolare tra gli studenti provenienti da contesti socioeconomici più svantaggiati o che ricevono l'istruzione in lingua albanese. I risultati delle analisi di regressione indicano che le aspettative educative e professionali, il background socioeconomico e la lingua di istruzione predicono in modo significativo le prestazioni in lettura. Le narrazioni degli studenti emerse dalle discussioni nei focus group rivelano inoltre che una limitata orientazione professionale, la sottovalutazione dell'impegno e il disallineamento tra aspettative e competenze rappresentano le principali sfide nello sviluppo di un'agenzia studentesca informata. Questi risultati di ricerca contribuiscono alla let-

teratura collegando l'agenzia degli studenti, la formazione delle aspettative e le competenze misurabili, offrendo indicazioni operative per la riforma educativa e l'orientamento professionale nella Macedonia del Nord.

1. Introduction

Adolescence is a critical period in which young people begin to imagine and evaluate their possible futures, they start forming expectations that influence their further educational choices, motivation, and longer-term life trajectories. Data from research studies suggest that these expectations do not emerge fully formed, but instead evolve gradually as adolescents gain experience and feedback from their social as well as institutional environments. So, early career-related thinking is often described as broad and idealised, shaped by cultural norms, family influences, and limited self-knowledge. Over time, increased exposure to educational demands and social comparisons contributes to more differentiated self-concepts and decision-making (Eccles *et al.*, 2003). According career development theories young people by the mid-adolescence begin to evaluate future goals in relation to their perceived abilities and available opportunities, although this process is uneven and strongly conditioned by contextual factors (Gottfredson, 1981, 1996).

Within the literature on career development, we can find the key analytical distinction between adolescents' aspirations and their expectations. Namely, aspirations describe the futures goals that young people would like to attain, reflecting their ideals, preferences, and socially valued outcomes, whereas expectations refer to the outcomes they believe are realistically achievable given their circumstances and perceived abilities (Gottfredson, 1981). This distinction is mainly relevant during mid-adolescence, when students begin to confront institutional constraints and performance feedback that may either reinforce or challenge their initial ambitions. According to Gottfredson's theory of circumscription and compromise, around the age of 14-15 young people enter a developmental phase in which occupational goals are increasingly filtered through assessments of feasibility, social accessibility, and self-evaluated competence (Gottfredson, 1996). However, evidence from PISA 2022 suggests that adolescents in North Macedonia hold very high educational and career expectations coexisting with weak performance in foundational academic domains, especially reading literacy. The country ranks 71st out of 81 participating systems in reading literacy, with 15-year-old students performing, on average, roughly six years below their international peers (Lameva and Andonova-Mitrevska, 2024).

This misalignment is analysed systematically by applying a mixed-methods design of quantitative and qualitative data. Data from descriptive statistics and focus group discussions are used to capture how adolescents articulate their educational and career expectations, how they perceive their own competencies, and which social or institutional factors they believe shape these views. At the same time, multivariate regression analysis is employed to test whether students' educational and career expectations, socioeconomic background and language of instruction are empirically associated with reading literacy performance. The quantitative and qualitative evidence are not treated as separate strands, but the study positions them as mutually informative. So, statistical patterns resulting from PISA 2022 data provide an objective benchmark against which students' narratives can be interpreted, while qualitative accounts offer context for understanding why certain expectations may be weakly aligned with academic achievement.

The research is grounded in the "expectations-values" model, which conceptualises individuals as inherently future-oriented, continuously predicting, expecting, and anticipating specific outcomes for their actions and behaviors. This anticipation serves as the driving force behind personal activity (Wigfield and Eccles, 2000). The model suggests that every motivation stems from an expectation of satisfaction, and since every prediction is based on prior experience, it follows logically that motivation includes a learning (cognitive) component. It also presumes certain cognitive abilities that allow individuals to envision future events (Wigfield, 1994). According to the "expectations-values" model, behaviour is more likely to occur if it is both valued and perceived as achievable within an individual's capabilities. Expectations reflect beliefs about one's likelihood of success in different tasks and activities, while values represent the motivations and reasons for engaging in a particular activity. This means that even if a person is confident in their ability to complete a task, they must still have a compelling reason to do so (Eccles and Wigfield, 2002). In Eccles and Wigfield's model, expectations and values are considered direct influences on achievement. These factors are shaped by task-specific beliefs, such as self-perceived competence, perceived difficulty of the task, personal goals, and self-concept. Furthermore, these social-cognitive variables are influenced by external factors, including others' expectations, interpretations of past performance, and affective memory (Wigfield and Eccles, 2000).

Research by Eccles and Wigfield within their "expectations-values" framework indicates a positive relationship between the valuation and expectation components. Students are more likely to place high value on tasks they believe they can perform well in, and vice versa (Eccles and Wigfield, 1995). According to the authors, this positive correlation helps maintain high self-esteem, as values and self-perceived abilities align rather than conflict. If an individual assigns high importance to a task they believe they are unlikely to complete successfully, frustration may result, leading to diminished self-esteem (Eccles and Wigfield, 2002). Applied to North Macedonia, this framework suggests that the coexistence of high educational and career expectations with very low reading literacy performance as documented in PISA 2022 reflects a potential misalignment between students' ambitions and academic preparation. Understanding this tension is crucial for interpreting students' aspiration not simply as individual preferences, but as outcomes shaped by both cognitive beliefs and structural conditions.

In addition to motivational beliefs, the concept of student agency provides an important lens for understanding how expectations are translated into educational and career-related action. Drawing on the OECD Future of Education and Skills 2030 framework, agency refers to young people's capacity to act intentionally, make informed choices, and shape their own life pathways through learning and development (OECD, 2019). Agency does not operate in isolation; it emerges through ongoing interaction with teachers, families, peers, and institutions. Constructs such as self-efficacy and beliefs about learning are closely connected to agency, as students who view ability as malleable are more likely to engage with academic challenges and persist in the face of difficulty. Exposure to information about education systems and labour market opportunities further supports the development of agency by enabling students to situate their aspirations within realistic pathways.

Student agency is not a personality trait; it is flexible and can be learned. Students require support from adults to exercise their agency and reach their full potential. From a young age, children begin to comprehend the intentions of those around them, fostering a sense of self, which is a crucial step toward developing agency (Woodward, 2009). As they advance through their education, students should be able to discover a sense of purpose in their lives and believe they can achieve that purpose by setting and pursuing goals. At this point, student agency becomes a key educational objective. The relationship between student agency and learning is reciprocal; when students actively participate in determining what and how they learn, they tend to exhibit increased motivation and are more likely to establish clear learning objectives. The development of agency is inherently relational, shaped by interactions with family, peers, and teachers over time (Schoon and Ng-Knight, 2017). This process is ongoing and evolves throughout an individual's life. Agency can be nurtured as students learn, receive feedback, and reflect on their experiences. The influences on a student's sense of agency include parents, peers, teachers, and the community, while students also impact the agency of their teachers, peers, and parents-creating a virtuous cycle that positively enhances children's development and well-being (Salmela-Aro, 2009). Thus, "co-agency", or "collaborative agency", highlights the role of an individual's environment in shaping their sense of agency. A productive learning environment is characterized by "co-agency", where students, teachers, parents, and the community collaborate effectively (Leadbeater and Dawson, 2017). One educational goal is to equip students with the necessary tools to realize their potential.

As education systems increasingly emphasize student agency, many young people nevertheless approach the transition to the labour market without sufficient resources or guidance to make informed decisions. Students whose career goals are ambitious, coherent, and aligned with their educational pathways are more likely to demonstrate the form of agency associated with successful school-to-work transitions (Covacevich *et al.*, 2021). Conversely, when aspirations are poorly informed or disconnected from academic preparation, this may signal underlying constraints that limit students' capacity to act effectively on their goals.

In traditional educational models, teachers are primarily responsible for transmitting knowledge through instruction and assessment. In contrast, systems that promote student agency emphasize learning as a process of co-construction, in which teachers and students act as partners. Within such frameworks, students develop a stronger sense of purpose and take greater responsibility for their learning, while teachers function as co-agents who actively support students' development (Calvert, 2016). In North Macedonia, however, the combination of ambitious educational and career expectations, low reading literacy performance, and limited exposure to labour market realities suggests that student agency is not yet fully realized. This gap highlights the need for interventions that simultaneously strengthen academic competencies and provide structured opportunities for informed career exploration, enabling students to align their aspirations more closely with their skills and opportunities.

2. Literature review

National longitudinal studies indicate that students' self-perception and outlook on their future occupation significantly impact their outcomes. Teenagers who feel uncertain, lack direction, have lower ambitions, or fail to recognize the relevance of education to their future are more likely to struggle in the labour market compared to their peers (Mann *et al.*, 2020). Often, young people's career expectations do not align with labour market demands. Furthermore, social background plays a key role in shaping these educational and career expectations, with students from lower socio-economic backgrounds and those with lower academic achievement demonstrating the most concerning patterns (Mann *et al.*, 2020). Career ambition, defined as the expectation of securing a managerial or professional job, has been linked to better employment outcomes in adulthood. These include higher wages, lower unemployment rates, and greater career satisfaction, even when accounting for background variables (Green *et al.*, 2023).

Data from PISA 2018 shows that a significant majority of young people expect to work in managerial or professional occupations by age 30, a trend that has been increasing since 2000 (Mann *et al.*, 2020). Research highlights the strong influence of socio-economic background and parental expectations on young people's educational and career expectations. For instance, longitudinal surveys reveal that while half of the children from managerial-class families follow their parents into similar roles, less than a quarter of children from manual worker backgrounds attain managerial positions (Agasisti *et al.*, 2024).

Findings from the *OECD Career Readiness Project* suggest that secondary school students who actively explore and reflect on their future careers tend to experience lower unemployment, earn higher wages, and find greater satisfaction in their careers as adults (Covacevich *et al.*, 2021). However, fewer studies have examined the specific barriers that children perceive as influencing their educational and career expectations. These barriers, whether real or perceived, can be numerous. For example, socio-economic status (SES) has been linked to occupation aspirations.

Children from lower SES backgrounds tend to have less knowledge about the job market, see fewer occupational options available to them, and aspire to roles that align with their SES level. As a result, their educational and career expectations are more limited compared to those of children from higher SES backgrounds (Hartung *et al.*, 2005). These findings suggest that educational and career expectations are socially patterned and may not equally translate into achievement.

An OECD study, *Challenging Social Inequality through Career Guidance* (OECD, 2024), demonstrates that young adults' labour market outcomes across OECD countries are strongly shaped by their social and economic backgrounds. Even with comparable education and skills, young adults whose parents did not complete upper secondary education are three and a half times more likely to be NEET (Not in Education, Employment, or Training) than those with at least one parent who completed tertiary education. Additionally, socially disadvantaged young adults are less likely to secure employment in service industries and are more often found working in fields such as construction and manufacturing. Even among individuals with equivalent education levels, those from higher socio-economic backgrounds are more likely to be employed in high-skilled professional sectors than their lower-SES peers.

PISA findings indicate that students often have a narrow view of career possibilities, with many focusing on a limited range of professions. Across OECD countries, at least half of 15-year-olds with defined job plans expect to work in just ten of the most common occupations among their peers (OECD, 2021). At an age when many students in OECD countries begin specializing in their studies, potentially limiting future career opportunities, uncertainty, confusion, and a heavy concentration of career expectations are widespread. PISA data also shows that students from lower economic, social, and cultural status (ESCS) backgrounds are less equipped to make informed decisions about their education and training. Longitudinal research indicates that such career expectations are linked to employment disadvantages in early adulthood (Covacevich *et al.*, 2021).

3. Method

3.1. Object and research hypothesis

The objective of the research is to examine the alignment between students' educational and career expectations, socioeconomic background, and language of instruction with their reading literacy achievement.

The hypothesis is: Students' educational and career expectations, socioeconomic background, and language of instruction are significantly associated with reading literacy. To explore this hypothesis a multiple regression model was used with all the variables (educational and career expectations, socioeconomic background, and language of instruction) entered simultaneously.

Focus group discussions were used to explore students' perspectives, on how they perceive the gaps between their expectations and their actual skills, and the barriers they face.

By integrating these methods statistical findings are interpreted alongside students' personal experiences, highlighting structural constraints and the role of student agency in shaping educational and career trajectories.

3.2. Sample

The sample is consisted of 111 secondary schools in North Macedonia, including 7,380 students who participated in PISA 2022. Of these, 6,610 students (approximately 90%) completed the reading assessment test. Assessments were administered in Macedonian (73%) and Albanian (27%) language (Lameva and Andonova-Mitrevska, 2024).

In addition to the quantitative testing, we conducted focus group discussions in eight secondary schools in Skopje, including four gymnasium and four vocational schools. Twelve students from each school participated, totalling 96 students.

4. Method(s) employed

4.1. Instrument and variables

Students completed computer-based test for assessment of reading literacy and a background questionnaire, which included measures of educational expectations, career expectations, and socioeconomic status (OECD, 2023).

Educational expectations were measured by asking students the highest level of education they expected to complete, coded according to ISCED levels (1-8) (OECD, 2023). For analytical purposes, three aggregated categories were used: primary education (ISCED 1), lower to post-secondary non-tertiary education ISCED 2 (ISCED 2–5), and tertiary education ISCED 3 (ISCED 6-8).

Career expectations were measured by asking students what kind of occupation they expected to have at age 30, coded using ISCO-08 major categories (0-9) (OECD, 2023).

Socioeconomic background, economic, social, and cultural status (ESCS index) was derived from parents' education and occupation, along with home possessions, and divided into low, medium, and high groups (OECD, 2023).

Language of instruction included two categories: Macedonian and Albanian.

Reading literacy achievements was defined as students' ability to understand, use, evaluate, reflect on, and engage with texts to achieve personal goals, develop knowledge, and participate effectively in society, including navigating both print and digital formats. Reading literacy achievement was measured using PISA 2022 plausible values, estimated through item response theory models based on students' responses to a subset of reading items (OECD, 2023).

Research included focus group discussions with students to explore their perceptions of the alignment between educational and career expectations and their actual competencies. Students were selected to represent diverse socioeconomic backgrounds and language-of-instruction contexts, ensuring variation in experiences and perspectives. Discussions were structured around key themes identified in the quantitative analysis, including students' awareness of academic requirements for different career paths, the role of parental and institutional guidance, and perceptions of learning support and assessment practices.

4.2. Statistical data analysis

Multiple linear regression analysis was performed using SPSS Statistics 21, with sample weights applied to ensure national representativeness. The categorical predictors, as educational and career expectations, socioeconomic background (ESCS), and language of instruction were dummy coded. The regression model examined the associations between these variables and students' reading literacy scores, providing insight into how educational and career expectations and background factors relate to academic performance.

The students' responses in focus group discussions were recorded, and analyzed using thematic analysis to identify patterns and recurring themes related to students' perceptions of educational and career alignment. Initial coding was conducted inductively, allowing themes to emerge from the data, followed by a deductive phase guided by key topics from the quantitative analysis.

5. Results

5.1. Descriptive statistics

5.1.1. Career expectations

Table 1 presents the ten most popular career expectations of students in North Macedonia alongside OECD averages. Overall, 57.4% of students reported specific career goals, slightly below the OECD average of 61.2%, while 71.7% aspire to one of the top ten occupations. ICT professionals are the most frequently cited (18.6%), followed by medical doctors (11.2%) and managers (8.1%), with other common choices including police and army officers (7.3%), legal professionals (4.9%), and sports and fitness workers (4.8%). Compared with OECD averages, we can see that students from North Macedonia tend to concentrate their expectations on a few high-demand professions, whereas OECD students' aspirations are more evenly distributed.

Tab. 1 – Career expectations of students (10 most popular expectations)

North Macedonia			*OECD*	
Rank	*Occupation*	*%*	*Occupation*	*%*
1	ICT professionals	18.6	Medical doctors	8.9
2	Medical doctors	11.2	Legal professionals	5.3
3	Managers	8.1	Architects, planners, surveyors and designers	5.0
4	Police and Army officers	7.3	Social and religious professionals	4.4
5	Legal professionals	4.9	Creative and performing artists	4.0
6	Sports and fitness workers	4.8	Engineering professionals (excluding electrotechnology)	4.0
7	Teachers	4.6	Sports and fitness workers	4.0
8	Architects	4.3	Other health professionals	3.8
9	Designers	4.3	Software and applications developers and analysts	3.7
10	Nursing and midwifery professionals	3.6	Nursing and midwifery professionals	3.4
	Total	71.7	Total	46.4

Table 2 shows managerial expectations across ESCS levels and language of instruction groups. Among high-ESCS students, managerial aspirations are higher among those receiving instruction in Macedonian (78.72%) compared to those receiving instruction in Albanian (21.28%). A similar pattern is observed among medium-ESCS students, with 68.75% for students instructed in Macedonian language than those students instructed in Albanian (31.25%). The same trend is observed among low-ESCS students, although the difference is smaller: students receiving instruction in Macedonian show a higher incidence of managerial expectations (52.93%) than their Albanian-instructed peers (47.07%).

We can see from these results that the likelihood of aspiring to managerial positions varies by socioeconomic background and is associated with the language of instruction, highlighting the interaction between ESCS and linguistic subpopulation in shaping career aspirations.

Tab. 2 – Percentage of students expecting to work in ISCO-08 major group 1 (Managers) by ESCS and language of instruction

ESCS level	*Macedonian*		*Albanian*		*Total*
	N	*%*	*N*	*%*	*N*
High ESCS	109	78.72	26	21.28	135
Medium ESCS	128	68.75	54	31.25	182
Low ESCS	31	52.93	23	47.07	54

A comparison of students' career expectations with national labour market needs reveals a clear mismatch. PISA data indicate that students' aspirations are concentrated in ISCO-08 major category 2 (professionals), while interest in skilled trades (ISCO category 7) is limited. However, according to the Employment Service Agency of North Macedonia (2023), priority for new employment over the next 12 months is given to skilled workers in ISCO categories 7-9.

5.2. Educational expectations

High percentage of adolescents in North Macedonia (71.8%) expects to pursue post-secondary education. Distribution by ISCED level shows 19.5% aim for a bachelor's degree (ISCED 6), 17.4% for a master's (ISCED 7), and 34.9% for a doctoral degree (ISCED 8). When we analyse the expectations by ESCS (Table 3), clear socioeconomic gradients emerge. Among high-ESCS students, the majority aspire to the highest educational level, with 85.20% expecting ISCED 3 (ISCED 6-8) qualifications, compared to only 13.82% expecting ISCED 2 (ISCED 2-5) and 0.98% expecting ISCED 1. Medium-ESCS students are more evenly distributed, with 68.85% expecting ISCED 3 (ISCED 6-8), 27.50% expecting ISCED 2 (ISCED 2-5), and 3.65% expecting ISCED 1. Students with low ESCS show lower aspirations: 57.16% expect ISCED 3 (ISCED 6-8), 38.96% ISCED 2 (ISCED 2-5), and 3.88% (ISCED 1). These patterns indicate that higher socioeconomic status is strongly associated with higher educational expectations.

Importantly, misalignment in this research does not imply that students from disadvantaged backgrounds necessarily hold higher expectations than their peers, but rather that their expectations often still highly ambitious in absolute terms are weakly aligned with their observed academic competencies.

Tab. 3 – Educational expectations of students by ESCS

	ISCED 1		*ISCED 2*		*ISCED 3*		*Total*
ESCS level	*N*	*% of ESCS*	*N*	*% of ESCS*	*N*	*% of ESCS*	*N*
High ESCS	16	0.98	201	13.82	1,234	85.20	1,451
Medium ESCS	75	3.65	577	27.50	1,434	68.85	2,086
Low ESCS	35	3.88	345	38.96	496	57.16	876

Note: ISCED 1 = primary education; ISCED 2 = lower secondary to post-secondary non-tertiary education (ISCED 2-5); ISCED 3 = tertiary education (ISCED 6-8).

Tab. 4 – Educational expectations of students by language of instruction

Language of instruction	*ISCED 1*		*ISCED 2*		*ISCED 3*		*Total*
	N	*%*	*N*	*%*	*N*	*%*	*N*
Albanian	42	3.92	315	28.33	735	67.75	1,092
Macedonian	84	2.47	808	24.24	2,429	73.29	3,321

Table 4 shows educational expectations by language of instruction. Students receiving instruction in Macedonian have higher incidence of aiming for ISCED 3 (ISCED 6-8) (73.29%) than students receiving instruction in Albanian (67.75%). In contrast, students receiving instruction in Albanian have slightly higher incidence of ISCED 2 (ISCED 2-5) (28.33% vs 24.24%) and ISCED 1 (3.92% vs 2.47%) compared to students receiving instruction in Macedonian language. These results highlight systematic differences in educational expectations associated with both socioeconomic background and language of instruction.

While students receiving instruction in Albanian report only moderately lower educational expectations than their Macedonian-instructed peers, their substantially lower reading literacy performance suggests a sharper misalignment between aspirations and academic preparation.

Based on the presented bivariate descriptive statistics we can see that there is a prevalence of ambitious career and educational expectations across different ESCS groups. The regression analysis further clarifies that these expectations are not neutral: some are systematically associated with higher achievement, while others are linked to lower literacy scores, indicating variation in students' informed agency.

5.3. Multiple linear regression analysis

The descriptive statistics show strong concentrations of ambitious educational and career expectations, but they do not indicate whether these expectations align with students' academic competencies. To examine this, a multiple linear regression was applied to test the extent to which students' educational and career expectations, socioeconomic background (ESCS), and language of instruction are associated with reading literacy scores.

The overall regression model was significant, F (14, 3885) = 123.31, $p < .001$, explaining 31% of the variance in reading literacy scores ($R^2 = .31$), indicating a moderate effect of educational and career expectations, socioeconomic status, and language of instruction on performance.

The intercept (CONSTANT = 407.69) represents the predicted reading literacy score for a student in the reference categories: receiving instruction in Macedonian language, expecting to complete tertiary education (ISCED 3, levels 6-8), having a high socioeconomic background (ESCS), and aspiring to a managerial career (ISCO 1). This value serves as the baseline against which the effects of all other variables in the model are interpreted (Table 5).

With regard to educational expectations, students expecting to complete lower levels of education achieved significantly lower reading scores compared to those expecting to complete tertiary education ISCED 3 (ISCED 6-8). Specifically, students expecting to complete primary education (ISCED 1) scored 34.76 points lower ($t = –5.83$, $p < .001$), while those expecting to complete lower secondary to post-secondary non-tertiary education ISCED 2 (ISCED 2 -5) scored 14.89 points lower ($t = -6.69$, $p < .001$).

Tab. 5 – The results from the multiple regression analysis

Variable	*B*	*SE B*	*β*	*t*	*p*
(CONSTANT)	407.69	2.43		167.75	
LI:ALB	-57.04	2.09	-0.38	-27.34	< .001
ISCED 1	-34.76	5.96	-0.08	-5.83	< .001
ISCED 2 (ISCED 2-5)	-14.89	2.23	-0.09	-6.69	< .001
ESCS low	-32.98	2.75	-0.20	-12.00	< .001
ESCS middle	-24.40	2.23	-0.18	-10.95	< .001
ISCO 2 Professionals	21.18	2.30	0.15	9.22	< .001
ISCO 3 Technicians & associate professionals	1.12	3.51	0.01	0.32	> .05
ISCO 4 Clerical support workers	9.84	9.06	0.02	1.09	> .05
ISCO 5 Service & sales workers	-23.25	3.71	-0.10	-6.27	< .001
ISCO 6 Skilled agricultural forestry & fishery workers	-75.80	26.89	-0.04	-2.82	< .001
ISCO 7 Craft-related trades workers	-30.65	5.47	-0.08	-5.61	< .001
ISCO 8 Plant &machine operators	-47.05	11.26	-0.06	-4.18	< .001
ISCO 9 Elementary occupations	-24.70	12.66	-0.03	-1.95	> .05
ISCO 0 Armed forces	-10.84	11.97	-0.01	-0.91	> .05

Note: N = 3,900.
Reference categories: Language of Instruction (LI) = Macedonian; International Standard Classification of Education (ISCED) = ISCED 3 (6-8); Index of Economic, Social, and Cultural Status (ESCS) = High; International Standard Classification of Occupations (ISCO) = Managers. B = unstandardized coefficient; SE B = standard error of B; β = standardized coefficient; t = t-statistic; p = significance.

Career expectations further differentiated reading literacy outcomes. Compared to students expecting to work as managers (reference category), students aspiring to professional occupations (ISCO 2) scored significantly higher (21.18 points; $t = 9.22$, $p < .001$). In contrast, students expecting to work in service and sales occupations (ISCO 5), skilled agricultural, forestry and fishery occupations (ISCO 6), craft and related trades (ISCO 7), and plant and machine operators (ISCO 8) scored significantly lower than those aspiring to managerial positions. Expectations to work in elementary occupations (ISCO 9), technicians and associate professionals (ISCO 3), clerical support workers (ISCO 4), and armed forces occupations (ISCO 0) were not statistically significant predictors of reading literacy achievement. Importantly, career expectations differentiate achievement even within the same educational system. Students aspiring to professional occupations outperform those aiming for service, craft, or machine-operating jobs, indicating that not all aspirations are equally grounded in academic competencies.

Students receiving instruction in Albanian scored, on average, 57.04 points lower in reading literacy than students receiving instruction in Macedonian ($t = -27.34$, $p < .001$).

Socioeconomic background also showed a strong association with reading literacy achievement. Students with low ESCS scored, on average, 32.98 points lower than students with high ESCS ($t = -12.00$, $p < .001$), while students with middle ESCS scored 24.40 points lower ($t = -10.95$, $p < .001$).

These results provide a quantitative benchmark against which students' perceptions and narratives, discussed in the following section, can be interpreted.

6. Discussion

The present research offers an integrated analysis of students' educational and career expectations and their academic competencies. The regression results demonstrate that educational expectations and selected career expectation categories, socioeconomic background, and language of instruction are significant predictors of reading literacy performance. Focus group discussions complement these findings by revealing how students themselves perceive the gap between their ambitious aspirations and their actual skills, frequently pointing to insufficient learning support, limited guidance, and socio-economic constraints. Importantly, the research reveals that the observed misalignment is not only the result of unrealistic individual ambition but also it reflects the interaction of personal beliefs, institutional practices, and structural inequalities. Moreover, not all career aspirations are equally aligned with academic competencies. Specifically, students who aspire to professional occupations demonstrate significantly stronger reading literacy outcomes than those aiming for service, craft, or machine-operating jobs. Although managerial careers are commonly perceived as prestigious, professional occupations such as ICT and healthcare typically require more direct and immediate academic preparation, which may explain their stronger alignment with reading literacy at age 15. This pattern indicates that career expectations differ not only in content but also in the extent to which they are grounded in measurable academic preparation.

Further, this research demonstrates how student agency is reflected in career expectations and whether these expectations align with academic performance, combining descriptive evidence from focus group discussions with predictive analysis of PISA 2022 reading literacy scores. For instance, many students, particularly those from lower socio-economic backgrounds or those receiving instruction in Albanian language, expressed high confidence in ambitious career plans despite limited awareness of the academic effort and competencies required to realize them. This misalignment is especially pronounced among students aspiring to lower-skilled occupations, whereas those targeting professional careers tend to display stronger alignment between aspirations and literacy performance. These differences suggest that structural factors shape not only achievement but also students' capacity to form informed and actionable expectations.

Students described an educational environment in which high grades are easily attained, sustained effort is undervalued, and assessment is perceived primarily as a summative judgment rather than a tool for learning. We can see that a central contributor to misalignment between aspirations and competencies is the weak recognition of the link between long-term academic effort and future career success. Many students reported prioritizing grades over deep learning and underestimating the level of preparation required for ambitious educational pathways. These perceptions echo PISA findings showing that students in North Macedonia place relatively little value on hard work as a determinant of success. Focus group discussions further revealed reliance on rote memorization and little engagement with cognitively demanding tasks, reinforcing inflated self-perceptions of academic readiness. As shown below, students' narratives consistently point to early assessment practices that reward compliance over competence:

> In primary schools, everyone gets grade "A" because the standards are very low. If you're slightly better than the others, teachers quickly encourage you by giving you the highest grade, and at home, your parents believe you're an excellent student.

Teachers in primary schools don't pay much attention; they leave us to our own. The focus is on minor details rather than the bigger picture, and everything revolves around grade "A", which is why grades shouldn't be given so much importance.

Our grading system is not as it should be. In elementary school, teachers tell you exactly which page to study from the textbook, and you can easily get an A. This is how we were conditioned from a young age-achieving the highest grade with minimal effort.

This is a good illustration of how early assessment practices may normalize minimal effort while simultaneously reinforcing unrealistic perceptions of academic competence. When students succeed with minimal cognitive effort, they receive weak signals about their actual skill levels, which may contribute to overconfidence and misaligned career expectations. This dynamic helps explain why some students' aspirations particularly those oriented toward lower-skilled occupations are weakly grounded in objective measures of literacy. By contrast, students aspiring to professional occupations tend to demonstrate stronger academic foundations, suggesting that alignment between aspirations and competencies is concentrated among more advantaged groups.

When we reflect on school context in North Macedonia, we can observe that sustained effort toward long-term educational and occupational goals is often insufficiently emphasized. There are not many opportunities for open discussion, reflection on learning goals, and feedback-driven improvement. The dominant perception of assessment as a final judgment undermines its formative function and reduces its capacity to support learning progression. Without formative feedback, students are deprived of the reflective processes necessary for developing informed agency. As a result of this, many students advance through the education system without developing a clear understanding of their own strengths and limitations. Nearly all students who pass the state matura exam, where pass rates exceed 90%, are eligible to enrol in university, with particularly high quotas for state-funded places (OECD, 2019). While this has increased the access to higher education, it has occurred alongside persistently low performance in international assessments, suggesting that many students enter tertiary education with substantial gaps in foundational knowledge and skills. These gaps hinder academic success and are compounded by misalignment between higher education programs and labour market needs (UNICEF, 2019). Consequently, recent university graduates in North Macedonia are less likely to secure employment than their peers in European Union countries (OECD, 2019), particularly in high-quality technical and vocational fields.

In research and policy recommendations on career development, career guidance is seen as a critical but underdeveloped mechanism for aligning aspirations with competencies. Participation in career development activities is consistently associated with improved employment outcomes, stronger motivation, and enhanced student agency (Cho and Ham, 2022). Although PISA 2022 indicators suggest that student participation in career guidance activities in North Macedonia is mostly comparable to OECD averages, focus group discussions point to important limitations in their depth and effectiveness. Many students reported that career-related discussions occurred too late, were superficial, or were absent altogether. They also reported that decisions regarding educational pathways were often shaped by parental expectations and anticipated earnings than by an informed exploration of interests, abilities, or labour market realities. Students aspiring to lower-skilled occupations appeared particularly affected by limited access to meaningful guidance, reinforcing patterns of misalignment between aspirations and academic preparedness. As we can see from the following narratives, students repeatedly described career decision-making as externally driven and weakly supported by schools:

In elementary school, we had a questionnaire to complete, and based on the results, I was advised to pursue something I had no interest in at all. So, I believe such questionnaires are pointless. We need open discussions when it comes to career counselling.

Students just sit and wait, doing nothing to work towards their adult goals. The curriculum is outdated, the classrooms are old, the lessons are boring, and there's no engagement.

Many of us aren't informed about where to go for secondary education. For me, the main support came from my parents. I'm still unsure if I made the right choice -I'll find out when I start my practical work in a medical clinic.

I think many students enter medical school because their parents push them to, and their ultimate goal is to finish school and move abroad for work.

The problem we face is that we're taught at home that only becoming a doctor or a programmer is considered success in life.

Students aim for higher education because they expect it to lead to higher earnings. Programming is a well-paid profession, both at the start and later in the career, which made it appealing to me.

These students' reflections indicate that career decision-making is often externally driven and insufficiently supported by structured, school-based guidance. Rather than facilitating informed choice, existing practices frequently reinforce parental expectations and narrow definitions of success. Although students are facing these challenges, they frequently reported high levels of confidence regarding their future trajectories. Nearly half of the students expressed concern about being unprepared for life after compulsory education and many reported experiencing family pressure in career decision-making, a majority nevertheless felt well prepared for their future and confident in their career-related decisions. This contradiction can suggest that confidence is not necessarily grounded in a realistic understanding of academic requirements or labour market conditions. Among students targeting lower-skilled occupations, high self-reported confidence may coexist with limited preparation, masking vulnerabilities that emerge during later educational and employment transitions.

Having in mind that schools represent an important opportunity to better align students' educational and career expectations with their abilities, the ongoing educational reforms in North Macedonia, including the introduction of a new gymnasium and vocational concept, should use the opportunity to embed career education more systematically within schooling. By supporting students in exploring career options, in developing their career management skills, and gaining authentic exposure to the world of work, these reforms can foster informed agency and co-agency. Ensuring that students' educational and career expectations are both realistic and achievable is essential for improving educational outcomes and facilitating smoother transitions into the labour market.

Reference

Agasisti T., Pastor J.M., Soler Á., Vicente I. (2024), "Career guidance to lessen inequalities in education: internal evidence from OECD-PISA 2018 data", *Education Economics*, pp. 1-18, https://doi.org/10.1080/09645292.2024.2343069.

Calvert L. (2016), *Moving from compliance to agency: What teachers need to make professional learning work*, Learning Forward & NCTAF, Oxford.

Cho Y., Ham S. H. (2022), "Does career guidance narrow the aspiration gap? Socioeconomic status and occupational aspirations of school children", *KEDI Journal of Educational Policy*, 19 (1), 45-66, retrieved on March 9, 2026, from https://www.kci.go.kr/kciportal/landing/article.kci?arti_id=ART002853904.

Covacevich C., Mann A., Besa F., Diaz J., Santos C. (2021), *Thinking about the future: Career readiness insights from national longitudinal surveys and from practice*, OECD Education Working Papers, No. 248, OECD Publishing, Paris, https://doi.org/10.1787/02a419de-en.

Eccles J., Templeton J., Barber B., Stone M. (2003), "Adolescence and emerging adulthood: The critical passage ways to adulthood", in M.H. Bornstein, L. Davidson, C.L.M. Keyes, K.A. Moore (*eds.*), *Well-being: Positive development across the life course*, Lawrence Erlbaum Associates Publishers, Mahwah, pp. 383-406.

Eccles J.S., Wigfield A. (1995), "In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs", *Personality and Social Psychology Bulletin*, 21 (3), pp. 215-225, https://psycnet.apa.org/doi/10.1177/0146167295213003.

Eccles J.S., Wigfield A. (2002), "Motivational beliefs, values, and goals", *Annual Review of Psychology*, 53 (1), pp. 109-132, https://doi.org/10.1146/annurev.psych.53.100901.135153.

Employment Service Agency of the Republic of North Macedonia (2023), *Survey on skills needs in the labor market in the Republic of North Macedonia for 2023, Results of the conducted research*, Employment Service Agency of the Republic of North Macedonia.

Gottfredson L.S. (1981), "Circumscription and compromise: A developmental theory of occupational aspirations", *Journal of Counseling Psychology*, 28 (6), pp. 545-579, https://doi.org/10.1037/0022-0167.28.6.545.

Gottfredson L.S. (1996), "Gottfredson's theory of circumscription and compromise", in D. Brown, L. Brooks (eds.), *Career*

Choice and Development, Jossey-Bass, San Francisco, 3rd ed., pp. 179-232.
Green S., Sanczyk A., Chambers C., Mraz M., Polly D. (2023), "College and career readiness: A literature synthesis", *Journal of Education*, 203 (1), pp. 222-229, https://doi.org/10.1177/002205742110022.
Hartung P.J., Porfeli E.J., Vondracek F.W. (2005), "Child vocational development: A review and reconsideration", *Journal of Vocational Behavior*, 66 (3), pp. 385-419, https://doi.org/10.1787/dfe0bf9c-en.
Lameva B., Andonova-Mitrevska T. (2024), *Report on Student Achievement in the Republic of North Macedonia [Electronic source]: PISA 2022 Programme for International Student Assessment*, State Examination Center, Skopje, retrieved on March 9, 2026, from https://dic.edu.mk/wp-content/uploads/2017/03/Izvestaj_NACIONALEN_PISA_2022.pdf.
Leadbeater E., Dawson E.H. (2017), "A social insect perspective on the evolution of social learning mechanisms", *Proceedings of the National Academy of Sciences*, 114 (30), pp. 7838-7845, https://doi.org/10.1073/pnas.1620744114.
Mann A., Denis V., Percy C. (2020), *Career ready? How schools can better prepare young people for working life in the era of COVID-19*, OECD Education Working Papers, No. 241, OECD Publishing, Paris, https://doi.org/10.1787/e1503534-en.
OECD (2019), *An OECD learning framework 2030*, Springer International Publishing, Cham, pp. 23-35.
OECD (2021), *Career Conversations: Why it is Important for Students to talk about Their Futures in Work with Teachers, Family and Friends*, OECD Publishing, Paris, https://doi.org/10.1787/15b83760-en.
OECD (2023), *PISA 2022 Assessment and Analytical Framework*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/dfe0 bf9c-en.
OECD (2024), *Challenging Social Inequality through Career Guidance: Insights from International Data and Practice*, OECD Publishing, Paris, https://doi.org/10.1787/619667e2-en.

Salmela-Aro K. (2009), "Personal goals and well-being during critical life transitions: The four C's-Channelling, choice, co-agency and compensation", *Advances in Life Course Research*, 14 (1-2), pp. 63-73, https://10.1016/j.alcr.2009.03.003.

Schoon I., Ng-Knight T. (2017), "Co-development of educational expectations and effort: their antecedents and role as predictors of academic success", *Research in Human Development*, 14 (2), pp. 161-176, https://doi.org/10.1080/15427609.2017.1305808.

UNICEF (2019), *OECD review of evaluation and assessment in education in North Macedonia: assessment and recommendation*, https://doi.org/10.1016/S0190-7409 (99)00071-7.

Wigfield A. (1994), "Expectancy-value theory of achievement motivation: A developmental perspective", *Educational Psychology Review*, 6, pp. 49-78, https://doi.org/10.1007/BF02209024.

Wigfield A., Eccles J.S. (2000), "Expectancy-value theory of achievement motivation", *Contemporary Educational Psychology*, 25 (1), pp. 68-81, https://doi.org/10.1006/ceps.1999.1015.

Woodward A.L. (2009), "Infants' grasp of others' intentions", *Current Directions in Psychological Science*, 18 (1), pp. 53-57, https://doi.org/10.1111/j.1467-8721.2009. 01605.x.

3. UNI.CO: a diachronic model for analyzing graduates' transition paths to employment

by Giulio Lucentini

This chapter presents the UNI.CO model, a diachronic approach for analyzing the transition from university to employment based on the integration of administrative data on graduates' careers and employment records (mandatory communications). Initially implemented through a collaboration between Sapienza University of Rome and the Italian Ministry of Labor, the model enables longitudinal tracking of employment outcomes for entire graduate cohorts in Lazio and Tuscany. By leveraging large-scale verified data, UNICO provides insights into contract types, durations, sectoral coherence, and geographic mobility. The model supports both regional policy design and university planning, offering indicators such as employment rates and coherence indices. The analysis highlights structural issues in the Italian labor market, including low contract stability and territorial imbalances, especially for graduates from southern regions and in STEM fields. These findings underline the importance of evidence-based strategies for improving both educational pathways and labor policies.

Il presente contributo illustra il modello UNICO, un approccio diacronico all'analisi della transizione dall'università al lavoro, basato sull'integrazione di dati amministra-

tivi relativi alle carriere dei laureati e alle comunicazioni obbligatorie del Ministero del Lavoro. Avviato grazie a una collaborazione tra la Sapienza Università di Roma e il Ministero del Lavoro, il modello consente di tracciare longitudinalmente i percorsi occupazionali dei laureati nel Lazio e in Toscana. L'utilizzo di dati certificati e di ampia scala permette di analizzare la tipologia e la durata dei contratti, la coerenza con il titolo di studio, e la mobilità territoriale. Il sistema supporta sia la programmazione universitaria sia le politiche attive del lavoro, attraverso indicatori come il tasso di occupazione e l'indice di coerenza. L'analisi evidenzia criticità strutturali del mercato del lavoro italiano, tra cui la bassa stabilità contrattuale e le disuguaglianze territoriali, in particolare per i laureati provenienti dal Mezzogiorno e per i percorsi STEM. I risultati sottolineano l'urgenza di strategie basate su dati per migliorare l'incontro tra formazione universitaria e domanda di lavoro.

1. Introduction

The transition from university to the labor market is a complex and often uncertain process, shaped by structural economic conditions, educational pathways, and regional labor policies. Understanding this transition requires analytical frameworks that go beyond simple employment rates, taking into account job stability, employment continuity, skill alignment, and career development over time.

The UNICO project was conceived to address these challenges through a data-driven and multidisciplinary approach to the analysis of graduates' employment trajectories. Launched in 2011 through a collaboration between Sapienza University of Rome and the Italian Ministry of Labor, UNICO integrates university administrative records with mandatory communications data, enabling the longitudinal observation of graduates' labor market participation. Over time, the project has progressively expanded its scope through research initiatives and institutional partnerships, incorporating national and regional labor market data and providing evidence-based insights for universities, policymakers, and students (Lucisano and De Luca, 2020).

The development of the UNICO model has unfolded through a series of strategic milestones that have strengthened both its methodological foundations and its territorial coverage. Between 2013 and 2016, the project was embedded within a national PRIN research program focused on educational success, social inclusion, and labor market integration, allowing for an in-depth analysis of career trajectories before, during, and after graduation (Lucisano *et al.*, 2016). In 2016, a scientific collaboration between Sapienza University and Luiss Guido Carli further reinforced the interdisciplinary character of the model.

A major step in the consolidation of UNICO was the creation of regional labor market observatories. In 2019, the Lazio Region Observatory was established through the collaboration of several universities (Cassino, LUISS, LUMSA, Roma Tre, Sapienza, and Tuscia) with the aim of systematically analyzing regional employment dynamics and supporting policy design (Alleva *et al.*, 2015). Building on this experience, in 2020 the Tuscany Region partnered with Sapienza to extend the model to universities across Tuscany, demonstrating its adaptability to different territorial contexts. By 2022, the regional observatories for Lazio and Tuscany were officially released, confirming UNICO's role as a structured analytical tool for labor market analysis and university planning.

More recently, in 2023, discussions were initiated with AlmaLaurea to explore the integration of the UNICO model with national graduate employment data, alongside the launch of a new observatory involving the universities of Triveneto. These developments represent a further step toward the creation of a scalable and replicable observatory framework.

Through this progressive evolution, UNICO has developed from an experimental initiative into a comprehensive diachronic model for analyzing transitions from higher education to employment. Its distinctive contribution lies in its ability to interpret employment outcomes as dynamic processes rather than static events, providing a robust empirical basis for understanding graduate employment pathways and for informing educational and labor policies.

2. Understanding the differences between the graduates registry and mandatory communications database

One of the key strengths of the UNICO project is its ability to integrate different data sources to offer a comprehensive view of how university graduates transition into the labor market. At the heart of this approach lie two fundamental databases: the graduates registry, which captures students' academic journeys, and the mandatory communications database, which records official employment data. While both are valuable on their own, they serve different purposes and have distinct limitations.

2.1. The graduates registry: a window into academic careers

The graduates registry is primarily maintained by universities and is designed to track students' academic progress from enrollment to graduation. It provides a detailed record of each student's educational background, including degree programs, courses taken, grades achieved, and final qualifications. This database is particularly useful for understanding the skills and knowledge graduates acquire before entering the job market.

However, when it comes to employment outcomes, the graduates registry has significant limitations. Universities often rely on graduate surveys to collect data on post-graduation employment, but these surveys are voluntary, conducted sporadically, and subject to response biases. As a result, the data may not fully represent the reality of graduates' career paths. Moreover, since the registry is education-focused, it lacks crucial labor market information, such as job types, contract stability, salary levels, and employer details.

2.2. The mandatory communications database: a snapshot of the labor market

Unlike the graduates registry, which is built around students' academic experiences, the mandatory communications database (comunicazioni obbligatorie – CO) is centered on the labor market. Managed by the Italian Ministry of Labor, this database records all formal employment contracts, as employers are legally required to report new hires, contract renewals, and terminations. This system provides a real-time and objective overview of labor market trends, making it a highly reliable source for tracking graduate employment.

The strength of the CO database lies in its completeness and accuracy. It offers detailed information on employment contracts, including the sector, company, location, salary range, and contract duration. It also provides insights into which industries are hiring, how long contracts last, and whether jobs are stable or temporary.

However, the CO database has its own limitations. While it effectively captures employment patterns, it does not contain educational data, meaning it cannot show whether a graduate's job aligns with their degree or whether their university background influenced their career choices. Moreover, it excludes self-employment, freelancing, and informal work, which are becoming increasingly relevant in today's economy.

2.3. How UNICO integrates the two databases

Individually, neither of the databases can fully capture the transition from higher education to employment. The graduates registry provides comprehensive data on students' academic backgrounds, including degrees earned, fields of study, and graduation dates, but it lacks any information regarding graduates' subsequent career trajectories or employment status. Conversely, the mandatory communications database accurately documents employment details such as job titles, sectors, and contractual arrangements but does not include any reference to individuals' educational qualifications or academic histories (Stanzione *et al.*, 2020).

UNICO effectively addresses this gap by integrating these two complementary datasets, allowing researchers and policymakers to construct a detailed and holistic picture of how educational pathways relate to labor market outcomes. This integrated approach enables a deeper understanding of how different university degrees prepare graduates for successful employment, identifying academic disciplines that offer the most stable and relevant employment opportunities. Furthermore, UNICO sheds light on the degree of alignment or mismatch between graduates' educational backgrounds and their actual employment roles.

3. Understanding graduate career pathways: a diachronic approach

Understanding the transition from higher education to employment requires a recognition of the complexities involved in career development. Graduates typically experience multiple stages in their employment journey, including initial job placements, short-term contracts, shifts in career direction, and sometimes periods of unemployment or further education. Acknowledging this, the UNICO project adopts a diachronic perspective, examining graduate careers over an extended period rather than providing only a static snapshot at a single point in time (Lucisano and De Luca, 2021). This diachronic approach is essential because it captures the full complexity and variability of career paths. Unlike traditional analyses, it highlights not just whether graduates find employment after completing their studies, but also how they secure their initial positions, how long they typically remain in temporary roles, when they transition to stable employment, and whether they experience significant shifts in their career trajectories. Furthermore, it enables detailed analyses of geographic and sectoral mobility, illuminating patterns related to regional movements, international migration, and industry transitions. Ultimately, by tracking these career pathways over time, UNICO provides policymakers and educational institutions with valuable insights that can inform targeted interventions and policies aimed at improving graduate employability and aligning higher education more effectively with labor market needs.

4. How UNICO reconstructs career paths

The ability to reconstruct career histories within the UNICO model derives from the integration of two key datasets:

– the graduates registry, based on university records, which provides information on graduates' degrees, fields of study, and academic performance;
– the mandatory communications database, managed by the Italian Ministry of Labor, which records all formal employment contracts, including type of employment, contract duration, employer, and location.

By linking these two sources at the individual level, UNICO is able to map graduates' actual employment trajectories, showing not only whether they found a job, but how their careers unfolded over time, through successive contracts, transitions, and possible interruptions.

A defining characteristic of this approach is its diachronic perspective. Employment outcomes are observed over the entire post-graduation period available within the 2008-2018 observation window, starting from the date of degree completion and extending to the end of the reference period. Consequently, the analysis does not focus on whether graduates enter employment within a predefined interval, such as one or two years after graduation, but rather examines how labor market participation develops over time.

This diachronic approach offers valuable insights for multiple stakeholders. For universities, it provides a more accurate evaluation of degree effectiveness, helping institutions adapt their programs to the realities of the labor market. For policymakers, it highlights structural issues such as contract fragmentation, prolonged employment instability, and mismatches between education and occupation, supporting the design of more targeted labor policies. For graduates, it offers a clearer and more realistic picture of early career trajectories, helping to contextualize expectations about labor market entry and progression.

The analysis of employment data for graduates from Sapienza University of Rome and Roma Tre University shows that, over the entire observation period, 62% of graduates, corresponding to 215,062 individuals, were matched with at least one employment contract recorded in the Mandatory Communications database. When restricting the analysis to contracts activated after graduation, the share decreases to 57%, corresponding to 199,390 graduates. This difference suggests that some graduates entered the labor market while still enrolled in university or were engaged in forms of work not fully captured by the administrative data, such as self-employment.

In total, graduates accumulated 1,013,215 employment contracts during the observation period, a figure that highlights the fragmented nature of many early career experiences. Rather than immediately securing long-term employment, many graduates moved through a sequence of short-term contracts, renewals, and job transitions. This pattern is also reflected in the total number of observed working days, which amounts to 301,728,434 days, corresponding to 63% of all observable post-graduation days. While these figures confirm sustained labor market participation, they also raise questions regarding job stability, contract duration, and the time required to reach more stable employment conditions.

Another important dimension of the analysis concerns geographic mobility. Graduates were employed in 5,552 Italian municipalities, covering approximately 70% of the national territory, as well as in 117 foreign countries. These patterns indicate a wide spatial distribution of employment opportunities, alongside substantial territorial differences in both job availability and job quality.

Overall, these findings show that graduates' entry into the labor market is not a single event but a process characterized by gradual adjustments, short-term engagements, and career shifts over time. While most graduates do enter employment at some point after graduation, the key challenge lies in understanding how and when stable employment is achieved. By reconstructing career paths over an extended period, UNICO provides a richer and more realistic representation of graduates' transitions from higher education to work.

5. Defining employment: understanding the employment index and graduate job placement

One of the key challenges in analyzing graduate employment concerns the definition of what it actually means to be "employed". In labor market analysis, a person is conventionally considered employed if they have worked at least one hour of paid work in a given reference week or have engaged in unpaid work in a family business. While this definition is consistent with international labor statistics, it does not adequately capture dimensions such as job stability, employment continuity, or income security, which are central to assessing the real impact of higher education on professional trajectories.

Within the UNICO framework, employment is therefore analyzed over an extended time horizon rather than at a single reference point. For each graduate cohort, employment outcomes are observed from the date of graduation up to the end of the available observation period, which in this study covers the years 2008 to 2018. This longitudinal perspective makes it possible to account for the cumulative nature of labor market participation in the years following graduation.

To provide a more meaningful measure of graduate employment, the UNICO project adopts the Employment Index. This indicator is calculated as the ratio between the total number of days spent in employment and the total number of days observable after graduation within the reference period, excluding overlaps between contracts. In this way, the Employment Index captures the intensity and continuity of labor market participation over time, rather than simply counting whether a graduate has held at least one contract.

This distinction is particularly important because many graduates experience multiple job transitions in the early stages of their careers, often moving through temporary or part-time positions before achieving more stable employment. By focusing on the accumulation of working days across the post-graduation period, the Employment Index provides a clearer picture of labor market integration, highlighting differences between continuous employment, fragmented trajectories, and periods of inactivity.

This methodological approach makes it possible to move beyond simplistic employment rates and to offer a more nuanced interpretation of graduate labor market outcomes. It also provides a consistent basis for comparing employment trajectories across cohorts and fields of study within the same observation window.

Empirical evidence from the UNICO data shows that a substantial share of employment contracts held by graduates are of very short duration. Specifically, 28% of contracts last only one day and a further 11% last no more than one week, pointing to widespread forms of employment instability. Although longer contracts, particularly those lasting between six and twelve months, may offer greater opportunities for skill development and professional learning, truly stable employment remains relatively rare, with only 6% of contracts extending beyond two years.

These findings highlight the limitations of traditional employment indicators that rely solely on job placement or contract type at a given moment in time. By incorporating a temporal dimension and focusing on job quality, stability, and career development over the entire post-graduation period, the Employment Index contributes to a more realistic assessment of how higher education translates into labor market outcomes.

6. Rethinking job insecurity: beyond traditional employment metrics

Traditional employment surveys tend to evaluate employment outcomes based on the type of job a graduate holds at the time of the survey, often equating stable employment with a permanent contract. While this approach provides a snapshot of the labor market, it fails to capture the fluid and often precarious nature of modern employment, where many graduates navigate through multiple short-term contracts and unstable work arrangements before securing long-term stability.

The data from UNICO highlights a trade-off phenomenon, where the most vulnerable groups, such as recent graduates, young workers, and those in certain sectors, are disproportionately affected by job insecurity. This suggests that relying solely on permanent contracts as an indicator of employment success is insufficient. Instead, a more comprehensive approach is needed, one that considers career trajectories over time, the frequency of contract renewals, and the extent of employment insecurity experienced across different job types.

By shifting the focus from static employment measures to a dynamic understanding of employment insecurity over time, it becomes possible to better address the challenges of labor market precarity and to develop policies that support not just access to employment, but more stable and sustainable career trajectories.

7. The role of the Coherence Index and Employment Index

Understanding graduate employment requires more than simply measuring whether a person has a job; it involves assessing how well their degree aligns with their professional path and whether they experience stable, continuous employment. Two key indicators help capture this complexity: the Employment Index, which measures the extent of workforce participation, and the Coherence Index, which evaluates how closely a graduate's job matches their field of study.

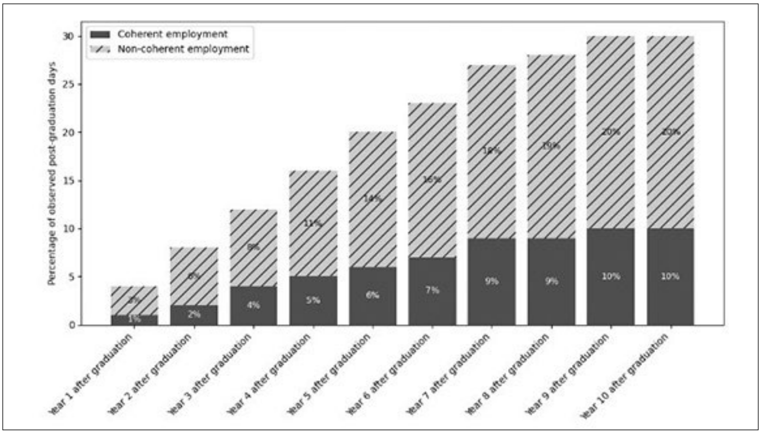

Fig. 1 – Employment and Coherence Index over the years

The Employment Index reflects the proportion of time a graduate spends in active employment relative to the total observation period. This metric moves beyond a simple employment/unemployment distinction, providing a clearer picture of job stability and continuity. It highlights whether graduates experience consistent work engagement or frequent periods of inactivity, offering a more nuanced understanding of labor market integration.

The Coherence Index, on the other hand, examines whether graduates are employed in fields relevant to their academic background. A low index suggests a weak link between tertiary education and the job market, indicating that many graduates work in roles unrelated to their studies. This disconnect raises critical questions about the effectiveness of academic training in preparing students for employment and whether job markets are recognizing university credentials as expected.

One striking insight from the data is that academic success, measured by high grades, does not show a strong positive association with stable employment outcomes. It is important to note that graduates who enrolled in doctoral or master's programs after graduation are retained in the dataset and classified as not employed during periods in which no employment contracts are recorded in the mandatory communications database. As a result, the observed relationship reflects differences in labor market participation rather than the exclusion of high-performing graduates who continue their education.

In some cases, a negative association emerges, indicating that graduates with higher grades do not necessarily secure more stable employment trajectories within the observed period. This finding challenges traditional assumptions about meritocratic returns in the labor market and suggests that factors beyond academic performance, such as work experience, employment opportunities, and structural labor market conditions, may play a more decisive role in shaping employment outcomes.

These findings call for a reconsideration of how universities and policymakers evaluate educational effectiveness. Instead of focusing solely on graduation rates and grades, a more comprehensive approach should integrate measures of job relevance, employment continuity, and career progression. By refining the way education and employment outcomes are assessed, it becomes possible to develop policies and educational strategies that better support graduates' long-term professional trajectories.

8. Balancing work and studies: the impact on academic and employment outcomes

A significant portion of university students work while pursuing their degrees, with 49% of three-year graduates having held at least one employment contract during their studies. While employment during university can provide valuable work experience and financial support, it also presents challenges. Data from the UNICO project suggests that working students tend to experience slower academic progression and often graduate with lower grades. Managing work and study commitments simultaneously can be particularly demanding, leading to delays in completing coursework and variations in academic performance across different fields of study.

At the same time, it is important to consider that employment during university may partially influence the relationship between academic performance and post-graduation employment outcomes discussed in the previous sections. Graduates who enter the labor market before completing their degree may display lower final grades while already being integrated into relatively stable employment trajectories. This dynamic can contribute to the observed weak or negative association between high grades and stable employment, without implying that academic achievement in itself reduces employment prospects.

Despite these academic difficulties, students who worked during their studies tend to be more active in the labor market after graduation. They often enter employment more quickly and accumulate greater work experience, even though their post-graduation employment is, on average, less closely aligned with their academic qualifications. This pattern suggests that early exposure to the labor market may facilitate employment continuity, while at the same time shaping career trajectories in ways that do not necessarily correspond to formal educational pathways.

Tab. 1 – Distribution of graduates working during and after degree completion

Degree	*Work*	*N*	*%*
Three year degree	Work during	34,421	48.8
	Work after	36,054	51.2
Master's degree	Work during	16,312	31.9
	Work after	34,850	68.1
Single cycle degree	Work during	4,639	41.7
	Work after	6,494	58.3

9. Employment outcomes of computer engineering graduates: job alignment and market dynamics

Analyzing employment patterns among Master's graduates in Computer Engineering from Sapienza University, we find that 90% are employed under subordinate or quasi-subordinate work contracts, indicating a high level of labor market participation. Over the observation period from 2008 to 2018, this group collectively accumulated 1,412,164 working days, highlighting their sustained presence in the workforce.

It is important to clarify that these results do not refer to employment outcomes measured at a fixed distance from graduation, such as one or two years after degree completion. Instead, employment and job alignment are observed over the entire post-graduation period available for each cohort within the 2008-2018 observation window. Consequently, graduates from earlier cohorts are observed for a longer period than more recent cohorts. The analysis therefore focuses on the overall structure of employment trajectories rather than on direct cohort-to-cohort comparisons at identical time horizons.

Within this framework, a closer examination of job alignment reveals a substantial degree of mismatch between educational qualifications and employment roles. While 55% of graduates are employed in positions aligned with their Master's degree in Computer Engineering, 45% hold jobs below their qualification level. Among these, 36.6% are employed in positions requiring only a Bachelor's degree, 56 individuals work as clerks, and 14 are employed in jobs characterized by very low qualification requirements.

Rather than simply confirming that a Master's degree in Computer Engineering facilitates access to employment, these findings draw attention to the persistence of underemployment even among highly trained graduates. The presence of significant mismatch between education and occupation highlights structural features of the labor market and points to limits in the translation of advanced technical skills into appropriately qualified positions.

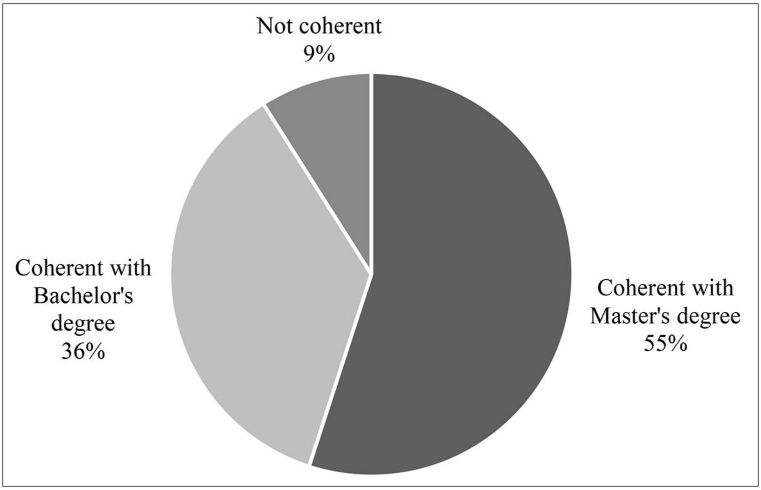

Fig. 2 – Coherence of total post-graduation working days for Computer Engineering graduates (Bachelor's and Master's degrees)

10. Rethinking university education and employment transitions

The transition from university to employment is undoubtedly influenced by education, but it should not be interpreted as the sole indicator of a university's effectiveness. Employment outcomes emerge from the interaction between educational pathways, individual strategies, and structural labor market conditions, which actively shape graduates' opportunities, expectations, and career trajectories.

One of the main challenges highlighted by this analysis concerns the narratives commonly used to describe graduate job prospects. These narratives are often based on partial or static indicators that fail to capture the complexity of employment trajectories over time. In the absence of dynamic and longitudinal analytical tools, decisions regarding educational choices and career planning risk being guided by simplified or outdated representations of the labor market. In this sense, a careful reading of observed employment patterns can contribute to avoiding the reproduction of past mismatches between education and work.

The results presented in this chapter underscore the importance of adopting an integrated analytical perspective that bridges education and employment. Indicators such as employment continuity, job-education coherence, and exposure to employment insecurity provide a more informative picture than point-in-time employment rates alone. From this perspective, the effectiveness of higher education cannot be assessed solely through graduation outcomes or academic performance, but should be examined in relation to graduates' longer-term employment trajectories and their capacity to access stable and meaningful work.

Several findings point to structural issues that merit attention. The persistence of fragmented employment paths, the limited association between academic achievement and stable employment, and the presence of significant job-education mismatch even among highly qualified graduates suggest that transitions to work are shaped by factors that extend beyond individual merit. Early labor market entry during university studies, sectoral characteristics, and territorial labor market conditions all contribute to shaping employment outcomes over time.

The UNICO model contributes to this debate by offering a diachronic framework capable of capturing these dynamics without reducing them to rankings or short-term performance indicators. By integrating administrative data on education and employment, the model supports a more cautious and evidence-based interpretation of graduate transitions, emphasizing processes rather than isolated outcomes. This approach also highlights the need for analytical tools that can accommodate emerging forms of work and non-linear career trajectories, moving beyond traditional definitions of employment and standard contractual arrangements.

From an educational perspective, this implies the need to rethink how universities, labor market institutions, and policymakers interpret and use employment data. Rather than focusing on immediate placement outcomes, greater attention should be devoted to employment continuity, job relevance, and the conditions under which graduates progressively build their professional paths. Addressing these dimensions requires time, reliable data, and interdisciplinary collaboration, avoiding simplistic judgments based on limited observation windows.

Ultimately, research on university-to-work transitions should maintain a critical stance toward dominant interpretations while preserving methodological rigor. Only through sustained collaboration between educational research, labor economics, and policy analysis can the complexity of graduates' employment trajectories be adequately understood and addressed.

References

Alleva G., Magni C., Lucisano P., Renda E., Petrarca F. (2015), *La domanda di lavoro per i laureati. I risultati dell'integrazione tra gli archivi amministrativi dell'Università Sapienza di Roma e del Ministero del Lavoro e delle Politiche Sociali*, Nuova Cultura, Roma.

De Luca A.M., Lucisano P. (2020), "Lavoro de generis. Gender at work", *Le società per la società: ricerca, scenari, emergenze. Atti del convegno internazionale SIRD, Roma 26-27 settembre 2019*, Pensa Multimedia, Lecce, pp. 156-167.

Lucisano P., De Luca A.M. (2020), "La transizione al lavoro dei laureati stranieri", in *Le società per la società: ricerca, scenari, emergenze. Atti del convegno internazionale SIRD, Roma 26-27 settembre 2019*, Pensa Multimedia, Lecce, pp. 188-200.

Lucisano P., De Luca A.M. (2021), *La carica dei 101. Storie di transizione al lavoro dei laureati stranieri*, Armando, Roma.

Lucisano P., De Luca A.M., Zanazzi S. (2017), "Educazione e transizione al lavoro. Strumenti per una migliore comprensione del fenomeno", in A. Notti (a cura di), *La funzione educativa della valutazione*, Pensa Multimedia, Lecce, pp. 647-664.

Lucisano P., Magni C., De Luca A.M., Renda E., Zanazzi S. (2017), "Percorsi di inserimento dei laureati nel mercato del lavoro attraverso l'uso delle Comunicazioni Obbligatorie (CO) del Ministero del Lavoro e delle Politiche Sociali", in L. Giovannini, I. Loiodice, P. Lucisano, A. Portera (a cura di), *Strategie orientative e transizione università-lavoro*, Armando, Roma, pp. 7-75.

Lucisano P., Magni C., De Luca A.M., Zanazzi S., Renda E. (2016), *Sapienza e lavoro. La domanda di lavoro e l'esperienza dei laureati*, Nuova Cultura, Roma.

Lucisano P., Renda E., Zanazzi S. (2017), "Stabilità lavorativa e alte qualifiche professionali. Uno sguardo sul fenomeno dell'overeducation a partire da fonti amministrative integrate", *Scuola democratica*, 8 (1), pp. 73-98.
Stanzione I., De Luca A.M., Poullain M., Lucisano P. (2020), "Costruire storie a partire da una lettura bottom-up dei dati amministrativi", *Lifelong Lifewide Learning*, 16 (37), pp. 58-72.

4. Machine learning models to identify students at risk of academic dropout

by Fiammetta Noccioli, Michele Marsili

This study aims to use supervised machine learning models, i.e., models trained on a set of features related to a known outcome, to explore their performance compared to more traditional models generally applied in education, analyzing the potential of this class of models that has seen increasing applications over the past few years. Specifically, the paper focuses on the prediction of university failure and its determinants by aiming to identify students who are at higher risk of dropping out of studies and to identify the most important predictors in this context. For the purpose of this study, it seems significant to predict this condition from the number of credits obtained at the end of the first year of the course. Indeed, relevant literature and the results of previous studies place emphasis on the transition from the first to the second year of the course as a turning point in the students' life and study paths. This study refers to a cohort of students graduated in the 2018/19 school year and enrolled at university in the 2019/2020 academic year, coming from a dataset built on the combination of different Italian data sources, Ministry of Education and Merit (MIM), University National Registry and INVALSI. Prediction of university failure was carried out using a set of machine learning techniques, known as supervised learning algorithms for classification problems. Several features, selected based on the literature and data availability and completeness, were evaluated in order to carry out the prediction, including individual-level characteristics, school type, and teacher grades. Infact, factors such as gender, age, educational career, influence the outcome of academic achievement, with high school final grade being particularly relevant here.

Machine learning can be an effective tool in predicting university success/failure and useful for orientation interventions and policy planning in the educational field, allowing students at risk to be identified early using the data available at the time of enrollment, and not career, and contextual factors.

In this context, INVALSI data introduce an extra wealth of information compared to other work predicting the phenomenon of university dropout.

Questo studio si propone di utilizzare modelli di machine learning con apprendimento supervisionato, ovvero modelli addestrati su un insieme di features relative a un risultato noto, per esplorare le loro prestazioni rispetto ai modelli più tradizionali generalmente applicati in ambito educativo, analizzando il potenziale di questa classe di modelli che ha visto un aumento delle applicazioni negli ultimi anni. Nello specifico, il lavoro si concentra sulla previsione del drop–out universitario e delle sue determinanti puntando a identificare gli studenti che sono a maggior rischio di abbandono degli studi e a identificare i predittori più importanti in questo contesto. A tal fine, sembra significativo prevedere questa condizione a partire dal numero di crediti ottenuti al termine del primo anno di corso. Infatti, la letteratura di riferimento e i risultati di studi precedenti pongono l'accento sul passaggio dal primo al secondo *anno di corso come punto di svolta nei percorsi di vita e di studio degli studenti.*

Questo studio si riferisce ad una coorte di studenti diplomati nell'anno scolastico 2018/19 e iscritti all'università nell'anno accademico 2019/20, provenienti da un dataset costruito sulla combinazione di diverse fonti di dati italiane, Ministero dell'Istruzione e del Merito (MIM), Registro universitario degli studenti e INVALSI. La previsione dell'insuccesso universitario è stata effettuata utilizzando un set di tecniche di apprendimento automatico note come algoritmi di apprendimento supervisionato per problemi di classificazione. Per effettuare la previsione sono state valutate diverse features, selezionate sulla base della letteratura e della disponibilità e completezza dei dati, incluse le informazioni di contesto, la tipologia di scuola e i voti degli insegnanti. Infatti, diversi fattori come il genere, l'età e il percorso di studi scolastico degli studenti hanno un impatto sull'esito del loro percorso accademico, con il voto del diploma di scuola superiore che risulta particolarmente rilevante in questo caso.

Il machine learning può rappresentare uno strumento efficace nel predire il successo/insuccesso universitario e utile per interventi di orientamento e pianificazione delle politiche in ambito educativo, consentendo di identificare precocemente gli studenti a rischio utilizzando i dati disponibili al momento dell'iscrizione e non solamente quelli di carriera.

In questo contesto, i dati INVALSI introducono una ricchezza di informazioni in più rispetto ad altri lavori di previsione del fenomeno della dispersione universitaria.

1. Introduction

Machine learning techniques, a branch of artificial intelligence, are increasingly being used in various fields of social research. Indeed, the increasing availability of data, including from administrative sources, makes it possible to use these techniques to identify recurrences and relationships between variables in different application areas. In the context of educational research, enormous progress has been made in recent years. One of the main trends is the use of predictive models aimed at detecting and preventing school and university dropout1 phenomena.

Several studies (such as Berens *et al.*, 2019) have used machine learning algorithms to try to predict, based on individual characteristics, students at risk of dropping out of university. However, data collection generally remains limited to the information level of individual universities (Serra, Perchinunno and Bilancia, 2017). The availability of data therefore remains limited to the university career and socio–demographic information, including only partial data on the student's secondary school performance. Therefore, data–driven analysis in higher education can contribute to the early identification of students at risk of failure (Yağcı, 2022). As in other studies investigating the relationship between high school backgrounds and university enrollment choices in Italy (Priulla *et al.*, 2024), the integration of different available data sources could make a greater contribution in this direction. This approach would

[1] Italy has shown significant improvements over the past ten years in terms of the percentage of graduates in the 25-34 age group (28.3% in 2021 compared to 21% in 2011). However, it still lags behind the average of European countries and the OECD average (ANVUR, 2023). The delay in tertiary education in Italy is due to various factors, including a high dropout rate (ANVUR, 2016).

allow, among other things, the inclusion of the skills measured in the INVALSI National Surveys as predictive factors of university failure.

2. Data

This study refers to a cohort of students graduated in the 2018/19 school year and enrolled at university in the 2019/20 academic year (248,847 students), coming from a dataset2 built on the combination of different data sources, Ministry of Education and Merit (MIM), University National Registry and INVALSI. The analysis is based on the information relating to each student enrolled in one of the Italian universities, which are present in the Registry at a bachelor's degree, with the exclusion of those who do not present information or implausible values for the variable number of credits at the end of the first year (around 5% of the sample) and students who decide to enroll at telematic university (5,3% of the sample).

Furthermore, are excluded:

– students who decide not to enroll at university the year after school;
– students who decide to enroll at university abroad.

For a dataset overview, the Figure 1 highlights with a Sankey diagram how students with different skill levels, based on their INVALSI scores in both italian and mathematics, proceed in their academic paths. The left side shows

[2] This study was conducted within the framework of the Agreement "From High School to job placement: An Analysis of University Careers and Student Mobility from Southern to Northern Italy" in collaboration with Italian Ministry of Education and Merit, Italian Ministry of University and Research and University of Palermo.

students classified as high-skill (black flows) and low-skill (grey flows), while the right side reports their enrollment outcomes by disciplinary area. The width of each flow is proportional to the number of students, thereby visualizing the relative distribution of enrollment choices across groups. In particular, around 76% of low-skill students (level 1 or 2) do not enroll in university the following year. In contrast, about 81% of high-skill students (level 4 or 5) continue to university, with 37% opting for STEM programs, followed by law, politics, and social sciences as the next popular area.

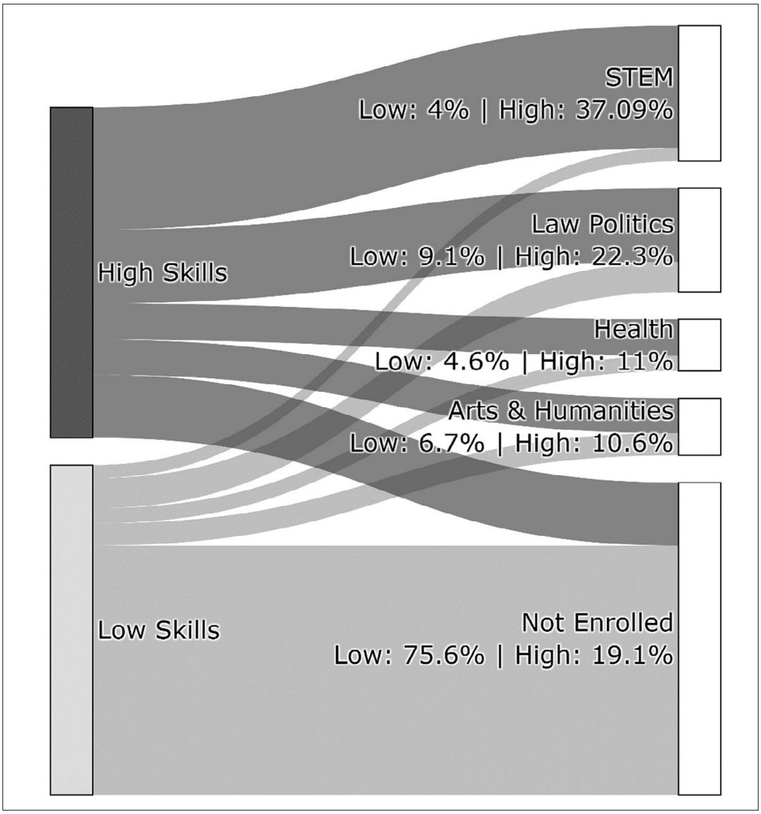

Fig. 1 – Sankey diagram of skills by enrollment categories

Tab. 1 – Variables overview

Variables	*Description*
Individual student information	
Reapiting student at grade 13	Students who have repeated at least one year during their school career or not
Migration background	3 categories: native, first generation immigrant students, second generation immigrant students
Sex	
Mother education	3 categories for mother education: pre-diploma, diploma, upper-diploma
Father education	3 categories for father education: pre-diploma, diploma, upper-diploma
School information	
School type	5 categories for type of school to which the students belong (lyceum, vocational, etc.)
Private/public school	Schools to which the students belong
Student performance at school	
High school final grade	Score between 60-100 for Matura exam at the end of school cycle
Results in the national surveys	
Italian grade 13 INVALSI score	Scores in the INVALSI Italian test at grade 13
Math Grade 13 INVALSI score	Scores in the INVALSI Math test at grade 13
University information	
Disciplinary area	4 macro categories of disciplinary area: Arts and Humanities, Law Politics, STEM, Health
Geographical macro-area	3 categories of the geographical area of the course attended by students: North, Centrum, South
Career events	Students have changing of curriculum, career, etc. in the academic year or not
Target variable	
Risk drop-out	Which equals one if there is a risk of University insuccess for the student career and zero otherwise

Our final dataset includes 205,905 university students and contains various types of information as shown in Table 1. It includes individual student details, such as if they are a repeating student at school, if they are a foreign student, their sex, and the education levels of their parents.

It also covers performance in secondary school and INVALSI tests in mathematics and Italian.

We have also university-related data like disciplinary areas, enrolments, and career events.

The target variable indicates if the student is at risk of drop-out or not at university, measured by the number of credits earned during the first year as outlined below.

2.1. Threshold definition for dropout risk prediction

For the purpose of this study, it seems significant to predict the risk of dropout from the number of credits obtained at the end of the first year of the course. Indeed, relevant literature and the results of previous studies place emphasis on the transition from the first to the second year of the course as a turning point in the students' life and study paths (De Angelis *et al.*, 2016; ANVUR, 2016). This critical issue is highlighted in the ANVUR's Biennial Report on the State of the University and Research System (ANVUR, 2023), which points out that the university dropout rate in Italy is among the highest in Europe. In our study we considered a threshold of credits as a proxy of academic insuccess, due to we don't have full data on this cohort's career outcomes.

To define the target variable, we tested different scenarios based on descriptive analysis of the dataset (quartiles, median, etc.), previous studies as benchmarks, clustering operations (K-means cluster) and results obtained from cohorts of students from previous academic years.

After a first approach with a single threshold for all students within the dataset, finally, we opted for thresholds derived from a Decision Tree tested on a cohort of students enrolled in 2016/2017 a.y for whom complete academic career and outcome data were available. In particular, we identified different thresholds for each academic area (4 categories) and we applied these to our cohort. These models showed an accuracy between 0.788 and 0.833 that confirmed the strong relationship between credits at first academic year and the risk of dropout. The thresholds obtained are 48 credits for "Arts, Humanities and Teaching"; 40 ones for "Economics, Law and Social Sciences"; 45 for "STEM" and, finally, 41 credits for "Health and Veterinary Sciences".

Therefore, it was taken into account that the numbers of credits required per year in Italian Universities is 60. Indeed, the standard duration of a Bachelor's degree program is three years and requires a student to obtain 180 credits in total.

We acknowledge the inherent limitation in calculating threshold for all students within the dataset, as the definition of a student being "at risk" can vary significantly depending on the specific context and policies of each university. The threshold for risk may be influenced by factors such as institutional requirements or regional academic standards. However, for the purposes of this preliminary study, our primary focus is on exploring the potential of machine learning techniques and identifying the most effective algorithms for predicting student risk, rather than determining an exact threshold for risk. This approach allows us to evaluate the potential relevance of machine learning models in this context.

3. Methods

Prediction of university failure was carried out using a set of machine learning techniques (ML), known as supervised learning algorithms for classification problems. Several features, selected on the basis of the literature and data availability and completeness, were used to carry out the prediction, including individual-level characteristics, and school type. Infact, factors such as gender, age, educational career, influence the outcome of academic achievement, with high school final grade being particularly relevant here.

Using the number of credits earned in the first year of university studies, considered a good proxy for dropout, the binary outcome variable was constructed as explained above.

Classification models used are logistic regression, a more traditional method, compared with machine learning techniques such as Naive Bayes, Gradient Boosting Machine and Random Forest.

Naive Bayes is an algorithm based on Bayes' Theorem, assuming conditional independence between features. It calculates the probability of each class and assigns an instance to the class with the highest probability. It is efficient, particularly suited for textual and categorical data (Zhang, 2004).

Gradient Boosting Machine (GBM) is an algorithm that sequentially builds decision trees, where each tree corrects the errors of the previous one by optimizing a loss function. It uses the gradient of the loss function to update model weights, improving predictive performance and reducing error. While highly effective on complex data, it can be prone to overfitting without proper hyperparameter tuning (Friedman, 2001).

Random Forest (RF) is based on an ensemble of decision trees, built on random subsets of data and features. It combines the predictions of individual trees using majority voting for classification, enhancing robustness and accuracy. Its structure helps reduce overfitting and effectively handles noisy and high-dimensional data (Breiman, 2001).

These methods were chosen as they are among the most relevant for solving classification problems according to the literature and for large dataset.

This approach allows us to learn from labelled data, improving the model's ability to make accurate predictions based on the features of each student.

The study is carried out following the stages of data pre-processing, model training and evaluation that are very common with ML models.

3.1. Data preprocessing

Before the analysis, an extensive data pre-processing phase was conducted. Infact, «transformations of the data to reduce the impact of data skewness or outliers can lead to significant improvements in performance» (Kuhn and Johnson, 2013, p. 27).

Firstly, the presence of outliers was checked by removing the last percentile of the distribution for the variable number of credits at the end of the first year.

Secondly, we standardize, recode and remove some variables to ensure comparability across features and improve model performance. Continuous variables were standardized to zero mean and unit variance, categorical variables were appropriately recoded using one-hot encoding and variables with redundant information were excluded.

Finally, the missing records identified on the context variables were treated. In particular, missing values (between 0.5% and 5%, except for the parental education around

15%) were replaced with a multiple imputation method and not a single one. The Multiple Imputation by Chained Equations (MICE) is a robust and informative method that use different predictive models (Azur *et al.*, 2011). In each iteration, each specified variable is imputed using the other. This process is continued until all specified variables have been imputed, generally in 5 iterations.

3.2. Model training, validation and evaluation

The dataset was divided into two parts: training and test sets, following a commonly used randomized distribution, respectively of 80% of the data set for the training and 20% for the test set, that provides a balance between reliable performance estimation and sufficient data for model training. Given the large size of the dataset, the use of this conventional split is reasonable.

We use a cross-validation (CV) technique to prevent overfitting and have robust estimate of the model's performance. Specifically, we chose a K-Fold CV procedure with k equal to 10 (Hastie, Tibshirani and Friedman, 2013). The dataset was split into 10 equal parts. The model is trained on 9 of these and tested on the remaining one. This process is repeated 10 times, each time using a different fold as the test set.

The performance of each model is measured using these well-known metrics: accuracy, precision, recall, F1 score and area under the ROC curve, where a high value corresponds to a better performance of the model (Sokolova and Lapalme, 2009). In particular, accuracy is the ratio of correctly predicted observations to the total observations. Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. It is also known as the positive predictive value. Recall, or sensitivity, measures the ratio of correctly predicted positive observations to all actual positives.

The F1 Score is the harmonic mean of precision and recall. It is a balance between the two metrics and is particularly useful when you need to take both false positives and false negatives into account.

Finally, the Area Under the Curve (AUC) represents the measure of the ability of the classifier to distinguish between the classes and is used as a summary of the ROC curve.

3.3. Advantages of Grid search

We tested also hyperparameter tuning technique to find the optimal hyperparameters for a model. This technique refers to the process of selecting the best set of hyperparameters for a machine learning model to improve its performance. Hyperparameters are parameters that are set before the training process begins and cannot be learned from the data itself. Common techniques for hyperparameter tuning include: Grid search, Random search and Bayesian Optimization.

In particular, we use the Grid search tuning technique which is very popular in ML context for its simplicity and power (Kuhn and Johnson, 2013). This algorithm uses a range of values for each hyperparameter and then evaluating the performance for all possible combination of these in order to find the best ones.

4. Results of machine learning models

The performance of the different models was evaluated using key classification metrics and the results highlight distinct trade-offs between accuracy, precision, recall, and AUC. We use a grid search technique for parameter selection as mentioned above. It can be seen from the Figure 2 that Logistic model shows a high recall, indicating that it successfully identifies most of the positive cases (at-risk students). However, the accuracy and precision are relatively lower (0.651 and 0.625). The RF model slightly improves both accuracy and precision (0.654 and 0.628) compared to Logistic Regression, indicating a better balance in classification. However, the GBM model outperforms both in almost all metrics except recall (respectively 0.659 for accuracy, 0.633 for precision, 0.740 for F1-score and 0.889 for recall), confirming its effectiveness in capturing underlying patterns in the data. Also, the AUC scores further support these results, showing that both models, RF and GBM, have similar discrimination capabilities between positive and negative classes (at-risk students and not at-risk students) even if with room for improvement (RF with an AUC of 0.744 and GBM with an AUC of 0.748).

Finally, The Naive Bayes model having the highest recall (0.912), which suggests that it rarely misses positive cases. However, this comes at the cost of a high false positive rate. This trade-off is crucial, as an overestimation of a-risk students might lead to unnecessary interventions.

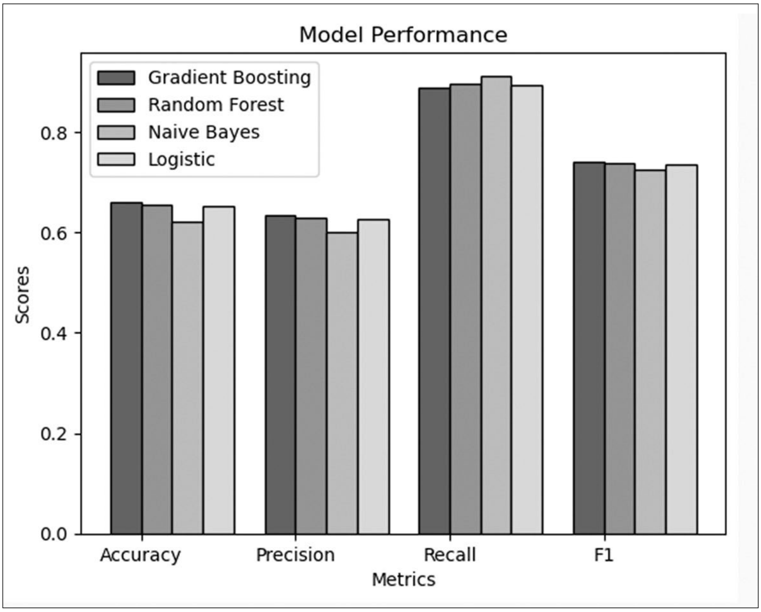

Fig. 2 – Comparison of different model performance

Source: Our model elaboration

In summary, the models show good sensitivity and quite good accuracy, with precision values to be improved.

The results obtained for these models indicate the gradient boosting model perform slightly better. This improvement is consistent with findings reported in other studies (Yağcı, 2022; Priulla *et al.*, 2024). Nonetheless, the very small performance differences observed suggest that predictive power is primarily constrained by the available features and data quality, rather than by the choice of model. This highlights the importance of data richness and feature relevance in driving improvements (Morán-Fernández *et al.*, 2022).

The levels of accuracy and precision could be attributed to the heterogeneity of the student population, as students come from diverse academic backgrounds, have varying levels of preparation, and face different personal conditions that influence their performance. This variability makes it challenging for the model to identify clear patterns and generalize predictions effectively, leading to reduced classification performance.

Nonetheless, these results provide a valuable starting point that individual universities can build upon by incorporating their own specific variables and contextual information. Such localized models are expected to further improve predictive accuracy and better address institution-specific factors, thus enhancing practical applicability.

Finally, the availability of new and updated data could make an important contribution to rethinking the target variable and testing additional models to improve predictive performance and practical applicability.

4.1. Feature importance

We also analyze the feature importance of the two most important models in terms of metrics: GBM and RF. In particular, Figure 3 shows the top 10 predictors in terms of importance for each model.

Each model uses distinct predictors, but some overlaps exist. *High school final grade* is the most important feature in both models. This result is consistent with previous studies conducted in Italy that have highlighted the strong relationship between high school final grade and university dropout (Atzeni *et al.*, 2022). The predictors *INVALSI scores*, *STEM disciplinary area* and *Geographical Area of studies* frequently appear across models, indicating their significance for the phenomenon. In particular, INVALSI scores in Math remain in second or third position across all 10 folds for both models, highlighting the importance of mathematical skills as a predictor of university failure risk. However, Italian skills also emerge as one of the most relevant factors, especially for the RF model.

The RF distributes importance more evenly, while the GBM focuses strongly on a few key variables. These differences arise because GBM tends to concentrate on the most predictive features, whereas RF spreads the weight more evenly across multiple features due to its ensemble nature.

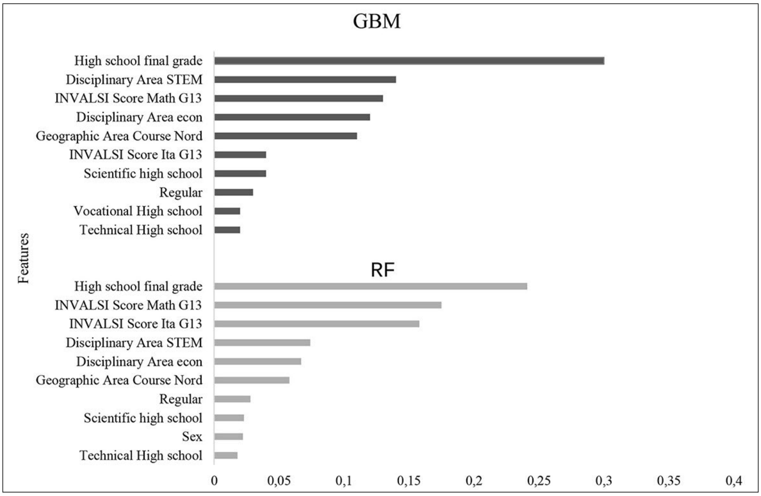

Fig. 3[3]*– Feature importance between Gradient Bosting and Random Forest*

5. Conclusions and future developments

In this preliminary study, we evaluated different machine learning models for predicting student dropout risk in

[3] Feature importance values range from 0 (no contribution) to 1 (maximum relative contribution).

a comparative analysis, identifying Gradient Boosting Machine as the best–performing approach. However, to further enhance predictive accuracy, we plan to test additional models such as XGBoost, CatBoost, and Artificial Neural Networks (ANN). From our analysis, the contribution of INVALSI data could also be valuable for the early identification of students at risk.

Our findings also led us to reflect on the factors influencing model performance and potential strategies for improvement. While we already employed cross–validation and ensemble methods such as Boosting techniques, additional refinements could be explored. For instance, incorporating new features – such as degree program information – might improve model robustness. Additionally, we have already tested hyperparameter tuning techniques, but further fine-tuning could yield better performance.

Future work will focus on integrating these enhancements while ensuring the model remains interpretable and practically applicable, aiming to be useful for refining dropout prediction and potentially support more targeted interventions and early strategies for students at risk of dropout.

References

ANVUR (2016), *Rapporto sul Sistema della formazione superiore e della ricerca 2016*, retrieved on March 9, 2026, from https://www.anvur.it/sites/default/files/2025-02/Rapporto_ANVUR_2016_integrale.pdf.

ANVUR (2023), *Rapporto sul Sistema della formazione superiore e della ricerca 2023*, retrieved on March 9, 2026, from https://www.anvur.it/sites/default/files/2024-12/Sintesi-Rapporto-ANVUR-2023.pdf.

Atzeni G., Deidda L.G., Delogu M., Paolini D. (2022), "Drop-out Decisions in a Cohort of Italian Universities", in D. Checchi, T. Jappelli, A. Uricchio (eds.), *Teaching, Research and Academic Careers: An Analysis of the Interrelations and Impact*, Springer International Publishing, Cham, pp. 71-103.

Azur M.J., Stuart E.A., Frangakis C., Leaf P.J. (2011), "Multiple imputation by chained equations: what is it and how does it work?", *Int. J. Methods Psychiatr Res.*, 20 (1), pp. 40-49.

Berens J., Oster S., Schneider K., Burghoff J. (2019), "Detection of Student at Risk – Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods", *Journal of Educational Data Mining*, 11 (3), pp. 1-41.

Breiman L. (2001), "Random Forests", *Machine Learning*, 45, pp. 5-32.

De Angelis I., Mariani V., Modena F., Montanaro P. (2016), "Immatricolazioni, percorsi accademici e mobilità degli studenti italiani", *Questioni di economia e finanza (Occasional Papers)*, n. 354 della Banca d'Italia.

Friedman J.H. (2001), "Greedy function approximation: A gradient boosting machine", *The Annals of Statistics*, 29 (5), pp. 1189-1232.

Hastie T., Tibshirani R., Friedman J.H. (2013), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, Berlin/Heidelberg.

Kuhn M., Johnson K. (2013), *Applied Predictive Modeling*, Springer, New York.

Morán-Fernández L., Bólon-Canedo V., Alonso-Betanzos A. (2022), "How important is data quality? Best classifiers vs best features", *Neurocomputing*, 470, pp. 365-375.

Priulla A., Albano A., D'Angelo N., Attanasio M. (2024), "A machine learning approach to predict university enrolment choices through students' high school background in Italy", *ArXiv abs/2403.13819*.

Serra A., Perchinunn P., Bilancia M. (2017), "Previsione dell'abbandono degli studi universitari mediante algoritmi di machine learning: un caso di studio su dati dell'Università degli

Studi di Bari”, in *Annali del Dipartimento Jonico*, vol. 5, pp. 396-421.

Sokolova M., Lapalme G. (2009), “A systematic analysis of performance measures for classification tasks”, *Information Processing & Management*, 45(4), pp. 427-437.

Yağcı M. (2022), *Educational data mining: prediction of students' academic performance using machine learning algorithms*, Smart Learn. Environ., 9, 11.

Zhang H. (2004), “The optimality of Naive Bayes”, *AAAI*, 1, pp. 562-567.

The authors

Clelia Cascella, before joining INVALSI as a researcher in 2014, she held a Ph.D. in Research Methodology and a PhD in Economics and then qualified as an Associate Professor in Mathematics and in Sociology. She has spent several years abroad, mostly in the UK, first as a Marie Curie Fellow and then as a Lecturer in Social Statistics.

Zaneta Chonteva, PhD, works at the Bureau of Education Development, Ministry of Education and Science, Republic of North Macedonia, as a advisor for research and research projects. Also, Chonteva is engaged as an external associate at the American University College-Skopje for the courses: Introduction to Statistics for Psychologists and Educational Psychology.

Beti Lameva, PhD is head of the Sector for Exams, IT and Research, National Examination Centre, Republic of North Macedonia. She is responsible for organizing state matura and national assessments, also she is the National Project Manager for PISA, TIMSS and TALIS studies. She has more than 20 years of professional experience in research in educational.

Francesca La Russa, she got her PhD in Foreign Language Teaching at Bordeaux Montaigne University and Roma Tre University. She works as a Researcher at INVALSI and is Associate member of the Laboratoire Cognition, Langues, Langage, Ergonomie (CLLE, UMR 5263) of Toulouse – Jean Jaurès University.

Giulio Lucentini, PhD candidate in the Piano Nazionale di Dottorato in Teaching and Learning Sciences, with a curriculum in Educational Research, at the University of Macerata and Sapienza University of Rome. His doctoral research focuses on the UNI.CO integrated data system for monitoring the school-to-work transition of graduates, in collaboration with IRPET (Tuscany). He is also engaged in the development of AI-based educational tools with locally trained models.

Michele Marsili, graduated in Statistical, Demographic and Economical Sciences at Sapienza University of Rome. He worked in Business Intelligence consulting, providing software development solutions for analysis and support for company's decision making in insurance and pharmaceutical industries. Since January 2018 he has been working in the INVALSI Statistical Department.

Fiammetta Noccioli, graduated in Statistical, Demographic and Economical Sciences at the University of Rome "La Sapienza". She obtained the II Level Master's degree in big data and machine learning. She previously worked as a statistician in epidemiological areas. Since November 2023 she has been working in the INVALSI Statistical Department.

In an increasingly complex educational landscape, access to and conscious use of data is key to understanding learning processes, reducing inequalities and supporting informed decision-making at instructional, institutional and policy levels. This volume, which collects a selection of contributions presented at the IX Seminar, "Data from and for educational system: tools for research and teaching" (Rome, 17–19 October 2024), explores how different data sources can contribute to a deeper understanding of how educational systems function, helping to improve students' educational and professional opportunities. Although the four chapters differ in terms of their methods, research questions and reference contexts, they are united by a common perspective: the importance of data as both an interpretative and an operational tool. Together, the four chapters offer a multifaceted perspective on the value of data in understanding educational phenomena, both in and out of the classroom. They demonstrate how various methodological approaches, such as psychometrics, multivariate statistical analysis, administrative data integration and predictive modelling, can converge towards the shared objective of developing more equitable, informed and responsive educational systems that better address the needs of students and society.

Patrizia Falzetti, Technologist Director, is the Head of the INVALSI Area of the Evaluation Research, of the SISTAN Statistical Office and of the INVALSI Statistical Service which manages data acquisition, analysis and return about both national and international surveys on learning (OECD and IEA). She coordinates and manages the process about returning data and statistical analysis to every school and to the Ministry of Education and Merit.

FrancoAngeli
La passione per le conoscenze