



MASTER THESIS

2025/2026

STUDENT NAME:

Dimitar DIMITROV

RESEARCH TOPIC

**AI in Education: A Catalyst for Learning or a
New Form of Dependency?**

- TUTOR'S NAME :

Prof. Ana Reyes Menéndez

The distribution of this document is strictly reserved to ESCP Europe.

CONFIDENTIAL

No

Table of Contents

Abstract:	3
Introduction	4
Chapter 1. Conceptual Framework, State of the Art and Research Design	5
1.1 Learning beyond performance: the distinction between deep and surface approaches.....	5
1.2 When the external tool short-circuits learning: cognitive offloading and dependency.....	7
1.3 Cognitive maturity of the learner as a determining condition	8
1.4 Generative AI in education: an unfinished empirical debate	11
1.5 Gaps in the literature	14
1.6 Working hypotheses	14
1.7 Research objectives	15
Chapter 2. Research Methodology	16
2.1 Population	16
2.2 Quantitative data collection: construction of the questionnaires	19
2.2.1 The Major Prépa survey as an external reference point	19
2.2.2 Construction of the questionnaire used in this thesis	21
2.3 Qualitative data collection: in-depth exchanges and observation	23
2.4 Data analysis	25
2.5 Methodological limitations	26
Chapter 3. Results	27
3.1 Sample description.....	27
3.2 AI use intensity and disciplinary distribution.....	27
3.3 Cognitive effort and perceived autonomy	31
3.4 Cognitive maturity.....	32

3.5 Self-reported outcomes and attitudes.....	34
3.6 Summary of findings	36
Chapter 4. Discussion, Conclusions and Managerial implications	37
4.1 Generative AI as a daily tool in constant evolution	37
4.2 Behavioural convergence but cognitive divergence between the two populations	39
4.3 The perceived gap between grades and durable understanding	41
4.4 The crisis of homework and students' knowledge and skills assessment: managerial recommendations	42
Bibliography	46

Abstract

The literature on generative artificial intelligence (AI) in education has converged on two parallel findings: significant short-term gains in academic performance and a measurable reduction in the cognitive effort that learners invest in their work. Most of the published studies treat student populations as homogeneous users, leaving the role of cognitive maturity, and the comparison between high school and selective French “classes préparatoires” tracks, poorly explored.

This thesis investigates the conditions under which generative AI in education supports deep learning rather than cognitive dependency. The study is based on a comparative analysis of two French populations exposed to the same technology at the same moment but at different stages of cognitive development: 42 students of classes préparatoires économiques et commerciales and 23 high school students. The 25-item questionnaire is built on the conceptual framework of deep learning, cognitive offloading, cognitive load theory, and metacognitive self-regulation.

The results show a behavioural convergence but cognitive divergence between the two populations. They use AI at similar frequency, in similar disciplines and for similar tasks, but differ significantly on every dimension of cognitive engagement. Prépa students more often try a task autonomously before consulting AI, accept its output less uncritically, preserve their capacity to work without it, detect errors more reliably and report a smaller decline in cognitive effort. The composite cognitive maturity index produces a gap of more than 1.5 points on a 7-point scale between the two populations, and the Spearman correlation between this index and the self-reported decline in effort since AI is negative and significant.

Both populations report that AI raises grades more than it deepens understanding. The thesis argues that the cognitive effect of AI is determined by the maturity of the learner rather than by the technology itself. Three converging managerial recommendations are derived: the redesign of homework around supervised, time-bounded and AI-free training conditions; the early training of students in the critical and reflective use of AI; and a dual assessment format combining unassisted examinations with AI-augmented oral defences that evaluate the reasoning trajectory.

Introduction

In November 2022, ChatGPT was released to the public. Almost no one in education saw what was coming. A tool that had been confined to a handful of research labs became available to anyone with an internet connection, and it could write essays, solve maths problems, summarise papers, translate texts and explain complex ideas in plain language. By the start of the 2023 academic year, surveys across European and American schools and universities showed that most students were already using generative AI for their academic work. Their teachers rarely knew and institutions had said nothing about it at that time.

No previous educational technology spread this fast. The calculator, the personal computer, the internet, the learning management system: all of them took years, sometimes decades, to settle into student life. Generative artificial intelligence (AI) did it in months. There was no institutional validation, no teacher training, no shared rule about what was acceptable and what was not. Students adopted the tool before schools had time to form an opinion. Teachers discovered the tool by grading papers, not by designing curricula. The usual order, where institutions introduce a technology and users gradually appropriate it, was reversed.

Schools and universities now find themselves in an uncomfortable position, finding it difficult to balance the pros and the cons. The pedagogical promise of these tools is important, as they give immediate, detailed feedback. They adapt to each student by making personalised tutoring available at a scale no human faculty could match. They let learners ask the questions they would never dare ask a professor, free of the fear of looking ignorant. For students who are struggling, who do not have the family resources to pay for private lessons, or who simply learn at a different pace, the democratising of the generative AI tools gives a real chance for progress. But the concerns are just as real. An AI can produce a finished assignment in seconds, hiding a complete lack of understanding behind a polished output. It can make it almost impossible for a teacher to tell whether a student has mastered the material or merely prompted an algorithm. And more deeply, it may end up weakening the very cognitive abilities that the educational system aims to develop, by removing the intellectual effort that has always been the price of learning.

The use of generative AI in education is the subject of an ongoing academic debate. Kasneci and colleagues (2023) argue that large language models can be powerful educational allies if they are used thoughtfully. Cognitive psychologists push back, warning of cognitive offloading

and dependency and what some now call cognitive atrophy or cognitive decline: when a mental operation is systematically delegated to an external tool, the underlying skill quietly erodes. It can aggravate the “brain rot” phenomenon, the cognitive decline and mental exhaustion experienced by individuals due to excessive exposure to low quality online material such as social media. Between these two camps, a third line of work, represented for example by Mollick and Mollick (2023), suggests that the effect of AI on learning is not determined by the technology itself but by how it is integrated into pedagogical practice. Most studies explored Anglo-American or Chinese undergraduates, which leaves selective and pen-and-paper examination systems such as the French classes préparatoires almost unstudied. Nevertheless, a survey of Major Prépa in May 2025 showed that 79% of preparatory class students are already using AI tools. This percentage means that classes préparatoires staff are obliged to re-examine their teaching based on the new reality. Moreover, most studies treat the users as a homogeneous group, although the cognitive and self-regulatory profile of an average high school student and that of a classe préparatoire student are not the same.

The **objective** of this thesis is to better understand under what conditions the use of generative AI in education stimulates deep learning among students, rather than a form of cognitive dependency, focusing on the French high school and classe préparatoire students.

Chapter 1. Conceptual Framework, State of the Art and Research Design

1.1 Learning beyond performance: the distinction between deep and surface approaches

Academic performance is not equal to learning. A correct answer in a test can be obtained through very different cognitive routes, and these routes do not have the same long-term consequences. The framework that allows the distinction to be made rigorously was introduced by Marton and Säljö (1976), in a study on how Swedish university students approached academic texts. Their central observation was that, faced with the same reading task, students adopted radically different strategies. Some focused on the literal content, on what they thought they would be asked to reproduce, and treated the text as an external object to be memorised. Others tried to grasp the meaning of the argument, to understand

what the author was trying to say, and to integrate this meaning into their own way of thinking. The first group displayed what Marton and Säljö (1976) called a surface approach, the second group a deep approach. The difference was not one of intelligence or effort but rather a difference of intention.

This distinction was later formalised by Biggs (1987) with the development of the Study Process Questionnaire, which allowed deep and surface approaches to be measured systematically across populations of students. Biggs (1987) showed that the two approaches are not stable personality traits but situated responses to the perceived demands of the academic environment. A student adopts a surface approach when the assessment system rewards reproduction, and a deep approach when the assessment system rewards understanding. Getting the right answer is, therefore, not the same as learning. A system that only grades the answer cannot tell whether the student understood or simply produced something that looked correct. Such a grading system cannot differentiate a student who made a genuine effort of learning from another who used an AI tool that can generate the right answer without thinking. This point is important for the argument of this thesis, because it implies that the form of learning that takes place in an institution is, to a large extent, a consequence of the form of evaluation that the institution practises.

Entwistle (1981), working in the same tradition, refined the framework by adding a third category, the strategic approach, which describes students who adapt their strategy to the demands of the moment, sometimes superficially, sometimes deeply, depending on what is at stake. The refinement matters here because it warns against any caricatural opposition between good and bad students. Most learners, in reality, oscillate between strategies. The relevant question is not who learns deeply, but under which conditions a given student adopts one approach rather than the other.

For the empirical part of this thesis, deep learning cannot be measured directly, since it refers to an internal cognitive process. The literature converges on three observable proxies. The first is the capacity to transfer knowledge to a new situation that was not part of the initial learning context. The second is the capacity to re-explain a concept in one's own words, with personal examples, without referring back to the source material. The third is the capacity to connect the concept to other concepts already mastered, through what cognitive psychology calls elaborative encoding. These three proxies guided the construction of the questionnaire

in part 2, and they are the criteria against which the effect of artificial intelligence on learning will ultimately be judged.

1.2 When the external tool short-circuits learning: cognitive offloading and dependency

AI tools can interfere with the cognitive mechanisms leading to cognitive dependency. Cognitive dependency is defined through three frameworks: cognitive offloading (Risko and Gilbert, 2016), the typology of cognitive loads (Sweller, 1988), and metacognitive self-regulation (Flavell, 1979; Zimmerman, 2002).

The first layer is the concept of cognitive offloading, developed by Risko and Gilbert (2016). Cognitive offloading refers to the use of physical action, or of an external device, to reduce the cognitive demands of a task. Writing down a phone number to avoid memorising it, using a calendar to keep track of appointments, asking a search engine for a piece of information rather than trying to recall it, all of these are forms of cognitive offloading. Cognitive offloading is, in itself, neither good nor bad. It is an adaptive behaviour that has accompanied human cognition since the invention of writing. The question is therefore not whether students offload cognitive operations to AI, but what kind of operations they offload, and with what consequences for the learning that the institution is supposed to deliver.

The second layer is the cognitive load theory developed by Sweller (1988), which provides exactly the conceptual tools to make that distinction. Sweller (1988) distinguishes three types of cognitive load. Intrinsic load corresponds to the inherent difficulty of the material being learned. Extraneous load corresponds to the additional difficulty imposed by the way the material is presented, by unnecessary complications, by poor instruction or by irrelevant distractions. Germane load corresponds to the cognitive effort that is directly invested in the construction of meaning, in the formation of mental schemas, in the integration of the new material with what is already known. Germane load is, in Sweller's framework, the cognitive cost of learning itself. It is what makes learning effortful, and it is precisely what produces the durable cognitive structures that the deep approach described by Marton and Säljö (1976) generates.

Offloading the extraneous load to an external AI tool is, therefore, desirable, since it frees mental resources for actual learning. Offloading the germane load is, on the contrary, deeply

problematic, because it removes the very effort that produces learning. The visible result, namely the correct answer to the exercise, is identical in both cases. The invisible result, namely the formation of the cognitive schema in the mind of the student, only takes place when the germane load is borne by the student themselves. This is the mechanism behind the AI paradox: AI can simultaneously improve performance but prevent learning.

The third layer is the work of Zimmerman (2002) on self-regulated learning, which is the capacity of a learner to plan their own activity, to monitor it during execution, and to evaluate it afterwards in order to adjust their future behaviour. This capacity rests on metacognition, that is, on the awareness that the student has of their own cognitive processes, a notion formalised earlier by Flavell (1979) and integrated by Zimmerman (2002) into a broader model of self-regulated academic behaviour. A student endowed with strong metacognitive abilities can recognise when an AI tool relieves them of the extraneous load, which is beneficial, and when it relieves them of the germane load, which is harmful. They can interrupt the use of the tool at the right moment, attempt the task on their own, compare their solution with that of the tool, and integrate the resulting feedback into learning. A student lacking these abilities, by contrast, tends to use the tool by default, without ever testing their own capacity, and progressively loses the very skills that the academic curriculum is supposed to develop.

Cognitive dependency on AI, therefore, is the state in which a student systematically offloads the germane cognitive load of their academic tasks to the tool, without metacognitive supervision, with the consequence that their autonomous capacity to perform these tasks gradually erodes. The frequency of AI use is not the relevant variable. The nature of the load offloaded, and the presence or absence of self-regulation, are the relevant variables. This definition is used in the questionnaire through a series of items inspired by what Risko and Gilbert (2016) implicitly suggest as a withdrawal test, namely the perceived capacity of the student to perform the task in the absence of the tool.

1.3 Cognitive maturity of the learner as a determining condition

The third concept needed for the argument is the cognitive maturity of the learner. It plays a central role because, as the next section will show, the empirical literature on AI in education produces contradictory results that cannot be resolved at the level of the technology itself. The same tool produces opposite effects on different students, and the moderating variable that explains this divergence is, in the framework of this thesis, the cognitive maturity of the

user. The definition of cognitive maturity is based on three complementary theoretical traditions.

The first tradition is that of developmental psychology, founded by Piaget (1964) on the stages of cognitive development. Piaget (1964) describes four successive stages through which the cognitive capacities of the child and the adolescent progressively unfold, namely the sensorimotor stage, the preoperational stage, the stage of concrete operations and the stage of formal operations. The passage from concrete to formal operations typically takes place during adolescence and is consolidated in early adulthood. During this passage the learner becomes capable of abstract reasoning, of hypothetical-deductive thought, of the manipulation of symbolic content independently of any concrete reference, and of the systematic evaluation of their own reasoning. These are the competences that the higher levels of academic work require, and they are also the competences that allow a learner to use an AI tool in a reflective rather than a substitutive way. A learner who has not yet consolidated formal operations tends to treat the output of an AI as a given to be accepted, because they do not yet possess the abstract evaluative framework that would allow them to question it. A learner who has consolidated formal operations can evaluate the output, compare it with their own reasoning and integrate it into a broader intellectual construction.

The second tradition is that of Vygotsky (1978) on the social nature of learning and on the zone of proximal development. Vygotsky (1978) defines the zone of proximal development as the space between what the learner can do alone and what they can do with the support of a more competent partner. Learning takes place through the progressive internalisation of the support received. AI can play the role of the more competent partner, but only under two conditions that echo the classical notion of scaffolding developed by Wood, Bruner and Ross (1976) in their elaboration of the Vygotskian framework. The first condition is that the support provided by the AI must be calibrated to the actual zone of the learner, meaning that it must address tasks that the learner could almost perform alone, not tasks that are entirely beyond their reach. The second condition is that the support must be progressively withdrawn as the learner internalises the competence, in a process of fading through which the scaffolding dissolves as the autonomous capacity emerges. For a cognitively mature learner, both conditions can be self-enforced, because the learner themselves monitors the calibration and initiates the fading. For a cognitively immature learner, both conditions fail, because the

learner does not possess the metacognitive competence to distinguish what they could do alone from what they cannot, and because the tool does not fade on its own. It remains permanently available, permanently willing to produce the output in place of the learner.

The third tradition is that of Zimmerman (2002) on self-regulated learning. If Piaget (1964) describes the structural maturation of cognitive abilities and Vygotsky (1978) describes the relational mechanism through which learning takes place, Zimmerman (2002) describes the observable competences through which cognitive maturity manifests itself in academic behaviour. These competences include planning, that is, the capacity to anticipate the cognitive demands of a task and to organise one's own activity accordingly; monitoring, that is, the capacity to track one's own progression during the execution of the task and to detect difficulties or errors as they arise; and evaluating, that is, the capacity to assess the quality of the outcome and to adjust one's future behaviour in consequence. These three competences are, according to Zimmerman (2002), the concrete manifestation of the metacognitive awareness that Flavell (1979) had identified as the foundation of mature academic cognition. They are also, in the context of this thesis, what distinguishes a student who uses AI reflectively from a student who uses it automatically. A planner knows when they need the tool and when they do not. A monitor detects the moment at which the tool begins to replace their own thinking. An evaluator compares the output of the tool with their own reasoning and identifies the points of agreement and disagreement. Without these three competences, the use of AI inevitably drifts toward the substitutive mode, regardless of the initial intention of the learner.

Cognitive maturity, therefore, designates the stable conjunction of the consolidation of abstract and hypothetical reasoning described by Piaget (1964), the capacity to interact productively with a more competent partner in the sense of Vygotsky (1978), and the metacognitive and self-regulatory competences formalised by Flavell (1979) and Zimmerman (2002). These three dimensions are developmental in nature, meaning that they are built progressively through education and experience, rather than being present or absent in a binary way. The cognitive maturity of a learner therefore exists on a continuum, and the relevant question is not whether a given learner is mature or immature in absolute terms, but where they stand on this continuum at the moment when they encounter the tool.

This continuum has a direct empirical implication. At the stage of secondary education, and

particularly during the first years of high school, the majority of learners are still in the process of consolidating formal operations, of developing their capacity to interact productively with adult knowledge and of acquiring the self-regulatory routines that abstract academic work requires. At the stage of higher education, and particularly in selective and competitive programmes such as the French classes préparatoires, the learners have, by construction, already reached a high level of cognitive maturity, because this level of maturity was the very criterion of their selection. The comparison between these two stages therefore offers a natural experiment on the effect of cognitive maturity, since the two populations are exposed to the same technology, in the same country, at the same moment in time, but differ structurally on the variable of interest.

It is important to note that cognitive maturity is not synonymous with age, nor with the level of study reached. A young learner can display remarkable maturity in a given domain, and an older learner can, on the contrary, regress toward a substitutive mode of use when confronted with a tool that covers a domain in which their own competence is weak. Cognitive maturity is a domain-sensitive variable, which protects the learner only on the domains where their own cognitive foundations are already strong. This nuance matters for the interpretation of the empirical results and for the formulation of the managerial recommendations, because it implies that the introduction of AI in education cannot be governed by a single threshold but must be modulated according to the specific cognitive profile of the learner on each domain of learning.

1.4 Generative AI in education: an unfinished empirical debate

The three preceding sections have constructed the conceptual lens through which the effect of AI on learning can be analysed. Before reviewing the literature, the perimeter of the technology under study must be defined. The expression AI covers an extremely heterogeneous family of tools, from learning management systems with adaptive features, to intelligent tutoring systems specialised in a given discipline, to large language models accessible through a conversational interface. The earlier review by Zawacki-Richter, Marín, Bond and Gouverneur (2019) identifies four broad families of educational AI applications, namely profiling and prediction, assessment and evaluation, adaptive systems and personalisation, and intelligent tutoring systems. The present thesis focuses exclusively on the last wave of this evolution, namely conversational generative AI, of which ChatGPT, released

in November 2022, is the most emblematic example, followed more recently by Claude, Gemini and others. This restriction is justified by two reasons. First, generative AI represents a qualitative break with the previous generations of educational technology, because it produces original content rather than retrieving or organising existing content. Second, it is the form of AI that students actually use on a daily basis, often outside any institutional framework, which makes it the more relevant object to study.

A first family of studies emphasises the educational potential of large language models. Kasneci and colleagues (2023) identify several pedagogical benefits, including the personalisation of explanations, the immediate availability of feedback, the lowering of the threshold of access to academic content for students with disabilities or non-native language backgrounds, and the support of teachers in tasks such as the design of exercises or the drafting of feedback. Mollick and Mollick (2023), in their working paper for the Wharton School, push this argument further by proposing a series of concrete pedagogical strategies through which AI can act as what they call a force multiplier for effective teaching, provided that the integration is intentional and pedagogically informed.

These early conceptual contributions have since been corroborated by results which, taken together, present consistent and measurable academic gains. The systematic review and meta-analysis conducted by Deng and colleagues (2025) on 69 studies published between 2022 and 2024 concludes that ChatGPT significantly improves academic performance, affective and motivational states, and propensity toward higher-order thinking, while at the same time reducing the mental effort that students invest in their academic work. Further meta-analytic work by Liu and colleagues (2025) reports a moderately positive effect of ChatGPT on academic achievement across 37 studies. The convergence of these meta-analyses, despite differences in their inclusion criteria and in the populations studied, supports a short-term improvement of academic performance. It has to be noted that Deng and colleagues (2025) reported that ChatGPT reduces the mental effort that students invest in their academic work, which is precisely the variable that section 1.2 identified as the cognitive cost of learning itself. A finding that is positive from the standpoint of short-term performance may therefore be ambiguous, or even negative, from the standpoint of deep learning as defined by Marton and Säljö (1976). Furthermore, the meta-analyses themselves warn that the durations of the experiments included are short, rarely exceeding a few weeks, which

leaves the question of long-term effects unresolved (Deng et al., 2025).

A second family of studies, equally numerous, draws attention to the risks. These risks are of three different natures, and it is important to distinguish them because they call for different responses. The first risk is academic integrity, namely the difficulty of distinguishing student work from AI-generated work, which calls for institutional and evaluative responses. The study by Cotton, Cotton and Shipway (2024) frames the problem as an integrity crisis that universities would have to manage through new policies, training and detection methods. More recent empirical evidence has shown that the phenomenon is large in scale: a Guardian investigation of 131 UK universities reported in June 2025 that confirmed AI-assisted cheating cases reached 5.1 per 1,000 students in the 2023-24 academic year, up from 1.6 per 1,000 the previous year, an approximately threefold increase (Lancaster, cited in The Guardian, 2025). The second risk is the propagation of errors and biases, since large language models produce text that is often plausible but not always accurate, a phenomenon now systematically described as hallucination (Ji et al., 2023). In the educational context specifically, hallucinations are particularly concerning because they undermine the trust on which any pedagogical relationship rests and can reinforce misconceptions rather than correct them (Danyaro et al., 2025), and the same logic extends to algorithmic bias, whose effects fall disproportionately on younger students and historically marginalised groups (Akgun and Greenhow, 2022). This risk calls less for institutional control than for the development of critical use skills among students themselves. Improvements of the AI models (such as Claude Opus 4.7) result in fewer hallucinations, and the technological development of the generative AI is likely to solve the problem of invented facts and false citations in the near future. The third risk, which is the one this thesis examines most directly, is the cognitive risk, that is, the possibility that the systematic use of AI weakens the cognitive abilities of students. The most direct empirical evidence on this point is provided by Gerlich (2025), who showed on a mixed-methods sample of 666 participants that frequent AI use correlates negatively with critical thinking scores on the Halpern Critical Thinking Assessment, with cognitive offloading acting as the mediating mechanism and the effect being most pronounced among younger users. Despite this study, the topic remains in its early development phase, and it suffers from a methodological difficulty, namely the impossibility of conducting long-term experiments at the scale required to observe a real cognitive evolution (Jose et al., 2025).

The contradiction between these two families of studies cannot be resolved at the level of the technology itself, because both families observe real phenomena. Some students, in some contexts, do learn more thanks to AI. Other students, in other contexts, do see their autonomous abilities decline. The resolution of the contradiction therefore requires the introduction of a moderating variable, capable of explaining why the same tool produces opposite effects in different contexts. This is the role played by cognitive maturity in the framework constructed above.

1.5 Gaps in the literature

Defining the optimal ways of AI use in education matters in practice for institutions, teachers, and students to maximise the positive aspects of it, while minimising the risk of cognitive decline. Institutions cannot simply totally ban the use of generative AI nor negligently tolerate it without control. They have to find a way to integrate it deliberately, developing ethical charters for its use, to preserve academic integrity, stimulate pedagogical innovation, and improve the actual quality of learning. Teachers have to rethink the design of their assignments, because an exercise that an AI solves in thirty seconds is no longer useful for the evaluation of students' knowledge and skills. Students have to learn when AI helps them progress and when it slows them down, because the immediate impression is often misleading. The stakes are particularly high in selective, competitive environments such as the French classes préparatoires aux grandes écoles, where the gap between a student who has truly mastered the material and one who has only produced correct answers can decide an entire trajectory. They are equally high earlier, at the high school level, where the basic cognitive habits of academic work are still being built, and where a poor use of AI can hamper cognitive abilities and prevent the students from learning how to learn.

1.6 Working hypotheses

The main working **hypothesis** of this thesis is that AI shapes student learning through a single mechanism, cognitive offloading, and the effects of this mechanism depend on what kind of load is being offloaded. When the load is extraneous, the student gains. When the load is germane, the student loses, because germane load is the cognitive effort that produces learning. Which of the two happens depends, in turn, on the cognitive maturity of the learner. A student with consolidated abstract reasoning, solid self-regulation and real metacognitive awareness can use AI reflectively and offload only what should be offloaded. A student

without these foundations tends to use AI as a substitute and offloads the very effort that would have built the foundations. The originality of this study lies in the comparison between classes préparatoires students and high school students. The prépa population, selected through a demanding academic track and trained to a high level of autonomous work, has, by construction, already built the cognitive foundations that allow a reflective use of the tool. The high school population is still in the process of building them. The two groups are exposed to the same technology, in the same country, appearing at the same moment in time, but they meet it at structurally different stages of cognitive development, which makes the comparison a natural test of the role played by maturity. The crisis revealed by AI is therefore not, at its root, a crisis of student behaviour. It is a crisis of validity in the assessment system itself which is not adapted to the new reality, and it calls for a redesign of academic evaluation around authentic and socially embedded tasks.

1.7 Research objectives

The **objective** of this thesis is to better understand under what conditions the use of generative AI in education stimulates deep learning among students, rather than a form of cognitive dependency, focusing on the French high school and classe préparatoire students.

To achieve this objective, the use of generative AI in two French educational populations that have been incompletely studied in the literature was documented: students of the classes préparatoires économiques et commerciales and high school students. The two populations were selected because they are expected to be enriched in learners with cognitive maturity already in place for the former and at a stage at which it is still under construction, for the latter.

At the managerial level, the objective is to derive a set of recommendations addressed to educational institutions that are currently trying to define their position on the use of AI. These recommendations are not framed as a list of best practices for the use of the tool, which would become obsolete with the next model release, but as a set of orientations for the redesign of assessment, on the principle that the cognitive question outlasts the model version.

Methodologically, the quantitative side relies on questionnaires distributed to students at the two stages of their educational trajectory, designed to measure the frequency and intensity of AI use, the perceived autonomy of the respondents, the cognitive effort invested in their

academic work, and their self-reported performance. The qualitative side relies on individual exchanges with students, conducted through in-depth written and voice conversations on social media for the older population, and through direct observation in an academic support context (home tutoring) for the younger one.

The detailed methodology is presented in chapter 2.

Chapter 2. Research Methodology

2.1 Population

The main case is the population of students in classes préparatoires économiques et commerciales, commonly called prépa ECG. The comparative terrain is a population of high school students reached through an academic support structure that provides home tutoring at the secondary level.

Since the French prépa system has no equivalent in most other countries, a brief description is needed before the methodological choices can be justified. The classes préparatoires aux grandes écoles are a post-baccalauréat track unique to the French system, in which students spend two years, sometimes three, in an intensive preparation programme aimed at the competitive entrance examinations of the grandes écoles, that is, the most selective business, engineering and humanities schools in the country. Admission to a prépa is itself selective, based on results obtained during the last years of high school. The prépa then functions as a second filter, through continuous evaluation and through the final national examination. The prépa ECG, the track that concerns this research, is the economics and business stream and prepares students for the entrance examinations of the top business schools. A student in prépa ECG typically works between fifty and seventy hours per week, follows a curriculum combining advanced mathematics, economics, history, geography, philosophy and two foreign languages, and is confronted on a weekly basis with written and oral examinations designed to simulate the conditions of the final concours. This description matters here because it establishes that a student in prépa ECG is, by construction, a cognitively mature learner in the sense given to that term in section 1.3, at least within the domains covered by

the curriculum.

The choice of the *prépa* ECG as the main case rests on three reasons, all directly connected to the conceptual framework of chapter 1. The first is the intensity of the cognitive workload. The volume of work demanded over two years of preparation is exceptional, which makes the question of cognitive effort particularly important. If students offload *germane* load to AI in this context, they are not offloading a marginal effort. The second reason is the selectivity of the final examination. At the time of writing, the national examinations that conclude the *prépa* are taken without access to AI. A student who can only perform with the tool is therefore a student unprepared for the actual evaluation, which creates a particularly sharp gap between performance with and without AI. The third reason is the historical emphasis of the *prépa* tradition on personal mastery of the material, which makes the arrival of AI particularly disruptive. Pedagogical formats such as the *khôlle*, the weekly composition and the oral interrogation place the student in situations where cognitive outsourcing is structurally impossible, and they contrast sharply with the home-work contexts in which AI use is, by contrast, entirely possible.

The comparative terrain is composed of high school students, reached through an academic support structure that provides home tutoring to more than one hundred pupils at the secondary level. The choice of this comparative terrain is useful, because it provides not only a comparison with another segment of higher education, but more importantly, a contrast with a stage at which the cognitive maturity is still under construction. High school students (*Lycée* in the French system), particularly in the final years, are confronted with cognitive tasks that resemble in form those of higher education, namely written compositions, structured arguments, mathematical problem solving and the reading of academic texts, which makes the comparison with the *prépa* meaningful. The second is the availability of the tool. High school students have autonomous access to a smartphone, to a personal computer and to generative AI tools, either through free versions or through subscriptions financed by their family, which means they can actually mobilise AI in their academic work independently of any institutional authorisation. The third is the developmental position of this population. High school students, typically aged between fifteen and eighteen, are in the phase that Piaget (1964) describes as the consolidation of formal operations, that is, at the moment when the cognitive foundations of abstract academic work are being built. If AI interferes with this

construction, the consequences are not marginal but structural, which gives the empirical investigation its particular weight.

The comparison between the prépa population and the high school population is not a comparison between good and bad students, nor between legitimate and illegitimate users of AI. The objective is not to judge one population against the other, but to observe how the relation between students and AI varies when the cognitive maturity of the learner varies, while other contextual variables are held as constant as the research design allows.

The academic support structure through which the high school population was reached is Prépa Rationnelle, a tutoring company founded by the author in 2024, which provides individual academic support at the secondary and post-secondary level. At the time of the survey, the structure followed more than one hundred pupils at the high school level, distributed across the three final years of secondary education, working on a weekly basis with a network of approximately fifty tutors recruited among current students of selective French Business Schools. The respondents were reached through this network, which gave the research two advantages that an external recruitment would not have provided. First, the relational trust built over months of weekly sessions improved the quality of the responses, since the pupils were answering through a channel they already associated with academic seriousness rather than with external research solicitation. Second, the tutors themselves were in a position to provide independent observations on the actual practices of their pupils. The author of the thesis follows within this support structure students in prépa ECG, who are among the students who replied to the survey. Again, this proximity and confidence built through the tutoring support structure Prépa Rationnelle allowed the collection of complete responses.

The sample is constructed as follows. For the quantitative component, the questionnaire was distributed through two channels adapted to each population. For the prépa population, the questionnaire was diffused on the Instagram account of the author, which has a large following among students engaged in prépa ECG studies. This channel, often overlooked in classical research designs, offered an advantage: direct access to a population that is notoriously hard to reach through institutional routes, because of the intensity of the workload and the relative closure of the prépa system to external research solicitations. Forty-two responses were collected on the initial screening question regarding AI use. For the high

school population, the questionnaire was diffused through the home tutoring structure mentioned above, which produced twenty-three complete responses, complemented by direct observations of the tutors working with these students on a weekly basis. The difference in sample size between the two populations is assumed and reflects the different conditions of access to each terrain.

For the qualitative component, twenty-two in-depth exchanges out of the 42 students that completed the survey were conducted with prépa students who had declared using AI, through voice and written messages on Instagram, and ten high school students followed by the academic support structure were interviewed in person on their relation to the tool. Several teachers of the structure also provided observations on the AI-related behaviour of their pupils, which enriched the qualitative material without constituting a separate data source.

2.2 Quantitative data collection: construction of the questionnaires

The questionnaire is the quantitative instrument of the research. Its construction follows directly from the conceptual framework of chapter 1, but it also draws on an external reference point, namely the survey conducted by Major Prépa in spring 2025 on the use of artificial intelligence by students in French classes préparatoires. It is not published in a peer-reviewed scientific paper, but is the closest that the author of this thesis found to the study question presented here.

2.2.1 The Major Prépa survey as an external reference point

In May 2025, Major Prépa, the principal media platform addressed to French prépa students, published the results of a national survey on the use of generative AI in classes préparatoires (Major Prépa, 2025). The survey is, to the best of our knowledge, the only relatively large-scale (n=440 participants, 74.8% of which in prépa ECG) empirical investigation conducted on this specific population to date. Its findings are widely cited in the French educational ecosystem and can serve as a benchmark for any conversation on AI in prépa. Aligning a part of the present questionnaire on the categories of the Major Prépa survey was therefore a deliberate methodological choice, enabling a direct comparison of the descriptive findings of this thesis with the national reference figures, on the variables that the two instruments share. The Major Prépa survey's exact questions and raw data are not publicly available to the

knowledge of the author, but from the analysis they presented, the following questions were deduced.

Category	Measured parameter	Example values or reported outputs
Sample characteristics	Number of respondents	
AI adoption	Use of AI tools	Yes / No
AI usage frequency	Frequency of AI use	Daily, several times per week, never, etc.
AI tools used	Type of AI used	ChatGPT, Gemini, DeepL, Google Translate, etc.
Academic disciplines	Subjects where AI is used	Languages, economics/HGG/ESH, culture générale, mathematics
Learning support	AI used to explain difficult concepts	Yes / No
Summarisation	AI used for chapter summaries	Yes / No
Information retrieval	AI used to search complementary information	Yes / No
Perceived pedagogical role	AI as ally or enemy	Attitudinal responses
Academic impact	Perceived impact on grades and results	Discussed qualitatively
Dependence perception	Concerns about overreliance on AI	Discussed qualitatively
Teacher replacement	Whether AI could replace teachers	Discussed qualitatively

Table 1. Reconstructed categories of the Major Prépa (2025) survey on AI use in classes préparatoires.

2.2.2 Construction of the questionnaire used in this thesis

The questionnaire developed for the present study integrates the descriptive categories of the Major Prépa instrument with constructs derived from the conceptual framework. It is organised in five thematic blocks: respondent profile, AI use intensity, cognitive effort and perceived autonomy, cognitive maturity, and self-reported outcomes and attitudes.

Block 1. Respondent profile

Construct: demographic and educational positioning. Items Q1 to Q4 capture year of study, academic track, age and gender.

Block 2. AI use intensity

Construct: AI use intensity. Items Q5 to Q7 aligned on the categories of the Major Prépa (2025) survey to allow comparison.

#	Item	Scale
Q5	On average, how often do you use a conversational AI tool (ChatGPT, Gemini, Claude, etc.) for your academic work?	Several times a day / Once a day / Several times a week / Once a week / Less than once a week / Never
Q6	Which AI tools do you use regularly for academic work? (select all that apply)	ChatGPT / Gemini / Claude / DeepL / Google Translate / Perplexity / Mistral / Other
Q7	In which subjects do you use AI? (select all that apply)	Maths / Languages / Economics or HGG-ESH or History-Geography / Culture générale or Philosophy / French / Sciences / Other
Q8	I use AI mainly to understand a concept I have not grasped from the course.	7-point Likert
Q9a	I use AI to generate ideas for essays or	7-point Likert

	dissertations.	
Q9b	I use AI for corrections or reformulations of my own writing.	7-point Likert
Q9c	I use AI to generate entire dissertations or written assignments.	7-point Likert
Q10	I use AI to summarise chapters or to retrieve information I could have found by myself.	7-point Likert

Block 3. Cognitive effort and perceived autonomy

Construct: cognitive dependency.

#	Item	Scale
Q11	When I face a difficult exercise, I first try to solve it by myself before consulting AI.	7-point Likert
Q12	When I use AI, I often accept its answer without checking whether it is correct.	7-point Likert (reverse-coded)
Q13	If AI suddenly disappeared, I could perform my academic work at the same level of quality.	7-point Likert
Q14	I can identify when an AI output contains an error or an inaccuracy in my discipline.	7-point Likert
Q15	Since I started using AI, I make less effort when I face an intellectually difficult task.	7-point Likert

Block 4. Cognitive maturity

Construct: self-regulatory and metacognitive competences.

#	Item	Scale
Q16	Before starting an academic task, I plan how I am going to approach it.	7-point Likert

Q17	While working, I regularly check whether I am still progressing toward my objective.	7-point Likert
Q18	After finishing a task, I evaluate the quality of my own work, independently of the grade I receive.	7-point Likert
Q19	When I face an unfamiliar problem, I can reason on it by myself without immediately looking for AI help.	7-point Likert

Block 5. Self-reported outcomes and attitudes

Construct: perceived effects on performance and on understanding, attitudinal positioning toward AI. Includes Likert-form reformulations of the Major Prépa (2025) attitudinal items.

#	Item	Scale
Q20	Using AI helps me obtain better grades.	7-point Likert
Q21	Using AI helps me genuinely understand what I am studying, in a way that I retain over time.	7-point Likert
Q22	I see AI more as an ally for my studies than as a threat to my training.	7-point Likert
Q23	I am worried about becoming too dependent on AI for my academic work.	7-point Likert
Q24	I am worried that AI use is weakening my reasoning abilities in the long term.	7-point Likert
Q25	Mastering AI tools will matter for my professional future.	7-point Likert

All items from Q8 onwards are measured on a seven-point Likert scale with the following anchors: 1 = Strongly disagree, 2 = Disagree, 3 = Somewhat disagree, 4 = Neither agree nor disagree, 5 = Somewhat agree, 6 = Agree, 7 = Strongly agree. The intermediate anchors are presented in full to the respondent in order to minimise the risk of central tendency bias.

2.3 Qualitative data collection: in-depth exchanges and observation

The qualitative component of this research took two different forms, adapted to each

population, in order to respect the specificities of each terrain while preserving the analytical coherence of the material collected.

For the prépa population, the qualitative material consists of twenty-two in-depth exchanges out of the 42 students who completed the survey, conducted through private messages on Instagram, in the form of voice messages or written messages. This format is unusual in academic research, but it produced a depth of exchange that a classical face-to-face interview would have struggled to replicate. The respondents were in their natural communication environment, they answered at the moment that suited them, and they did so with a degree of spontaneity and honesty that an institutional interview setting does not always generate. The exchanges covered the core dimensions of the interview guide built during the methodological phase: the types of tasks for which AI is used, the moment at which the student turns to the tool, the perceived effect on understanding, the capacity to perform the same task without AI, and the general attitude toward the tool. The written and voice format allowed each respondent to formulate their answers at their own pace, which is a real advantage in a population characterised by extreme workload and limited availability for synchronous interactions.

For the high school population, the qualitative material consists of individual exchanges with ten students followed through the home tutoring structure, complemented by observations gathered by the tutors themselves during their weekly sessions and by feedback from several teachers of these students. The exchanges with the pupils were more directed than those conducted with the prépa population, both because the age of the respondents called for a more guided conversational setting and because the home tutoring context provided a regular and trustworthy relational frame. The observations gathered by the tutors are particularly valuable here, because they provide an external perspective on the actual practices of the students, as opposed to the self-reported perspective of the students themselves. The divergence that sometimes appears between what a student declares about their use of AI and what their tutor actually observes in the session is not an inconvenience but a finding, which will be exploited in the analysis.

All exchanges and observations were conducted in French, since both populations are francophone, and the material was anonymised before analysis. Written exchanges were preserved in their original form, voice messages were transcribed verbatim, and tutor

observations were consolidated into structured memos after each session. The translation of the quoted extracts into English, for the final thesis, was performed in a way that respects the idiomatic expression of the respondent while producing a readable English text, with the original French version kept in the research file for verification.

The selection of the respondents followed a logic of maximum variation within each population, in order to ensure that the qualitative material captures the diversity of profiles rather than a single dominant type. Within the *prépa* population, the sample includes students from first year and second year, with different levels of academic performance and different self-declared intensities of AI use, as well as a group of students who declare not using AI at all, whose articulated reasons for abstention constitute a finding in their own right. Within the high school population, the sample includes students from different grades and with different academic trajectories, ranging from students who use AI systematically to students who use it only occasionally.

2.4 Data analysis

The initial analysis of the data was to generate the descriptive statistics. For each item, the mean, the standard deviation and the response distribution are computed separately for the *prépa* population and for the high school population. The descriptive output produces the profile of each sample and identifies the items on which the responses are concentrated, dispersed or polarised. For Likert items, mean scores are read on the 1 to 7 scale: scores above 5 indicate clear agreement, scores below 3 clear disagreement, and scores between 3 and 5 a more nuanced position that may justify a closer qualitative reading.

Next, the results were compared between the *prépa* population and the high school population. For each Likert item, the difference between the two means is tested with a Mann-Whitney test (two-sided), as due to the small sample sizes, the assumption of normality is not verified for most of the parameters. The comparison of proportions between the *prépa* population and the high school population of the present study, and between the *prépa* sample and the Major Prépa (2025) national distribution, is conducted using Fisher's exact test. The conventional significance thresholds are applied: $p < 0.001$ (***), $p < 0.01$ (**), $p < 0.05$ (*), $p < 0.10$ reported as a trend (+).

Finally, the relations between the parameters were examined, including: the correlation

between AI use intensity (Q5) and self-reported outcomes (Q20 versus Q21), in order to test whether intensive users report better grades but lower durable understanding; the correlation between the cognitive maturity index (Q16 to Q19 averaged) and the cognitive dependency items (Q11, Q12 reverse-coded, Q13, Q14, Q15); and the divergence between Q20 (better grades) and Q21 (durable understanding) at the individual level. No causal claim is drawn from these correlations: the analysis describes associations, not causal effects. The Python code used for the data analyses and for figure generation was verified by Claude Code. The findings of this quantitative analysis are then brought together with the qualitative material, when each conclusion is examined against both data sources. Where the two sources point in the same direction, the confidence in the finding is reinforced.

2.5 Methodological limitations

Several limitations of the methodological design have to be taken into account.

Self-reporting and social desirability. All Likert items measure what the student declares about their own practice, not what they actually do. Two effects compound this: students who use AI intensively in contexts where it is discouraged may tend to underreport, and students who value autonomy may tend to overestimate how much of their work they do without the tool. The questionnaire mitigates this through indirect formulations, neutral framing and full anonymity, but the bias cannot be eliminated. The comparison between the study populations is less affected, since the bias is likely to operate in the same direction in both groups.

Sample size and recruitment bias. The high school subsample is small ($n = 23$), which limits the statistical power of the within-group analyses and rules out fine-grained subgroup comparisons. Recruitment channels add a second layer of bias. The prépa questionnaire was diffused through an Instagram account, which over-represents students active on that platform and among the students of Prépa Rationnelle. The high school questionnaire was diffused through Prépa Rationnelle. Prépa Rationnelle is a tutoring structure of which the author is the founder, which over-represents families that invest in academic support. Therefore, the students may be better educated on how to use AI optimally and may have access to paid AI accounts. The diffusion of the questionnaires through Prépa Rationnelle may also create a conflict of interest. The bias is mitigated by the anonymity of the questionnaire, by the absence of any institutional pressure on the families to participate and by the fact that

the analysis does not concern the quality of the tutoring service itself.

Cross-sectional design. The data captures a single moment in time. The analysis can describe associations between AI use and cognitive autonomy at a given point, but it cannot establish whether intensive use erodes autonomy over time, or whether students with lower autonomy turn more often to AI. A longitudinal design would be needed to disentangle direction of causation, and this is beyond the scope of a master's thesis.

Chapter 3. Results

3.1 Sample description

The quantitative analysis includes 42 respondents from the prépa ECG population and 23 respondents from the high school population. Within the prépa sample, all respondents are enrolled in the economics and business stream; the distribution between first and second year is balanced. Within the high school sample, the respondents span the three final years of secondary education. Both samples are within the age range expected for their respective educational stage, and the distribution by gender is balanced in each. The descriptive statistics presented in the following sections compare these two populations item by item.

3.2 AI use intensity and disciplinary distribution

All 42 prépa respondents and all 23 high school respondents declare using a conversational AI tool for their academic work. The screening question on AI adoption therefore produces a 100% positive response in both populations, which mirrors the trajectory observed in the recent French and international surveys reviewed in chapter 1. Table 2 reports the frequency distribution.

Frequency	Prépa ECG	High School
Several times a day	6 (14.3%)	1 (4.3%)
Once a day	7 (16.7%)	5 (21.7%)
Several times a week	23 (54.8%)	13 (56.5%)
Once a week	6 (14.3%)	4 (17.4%)

Table 2. Frequency of AI use for academic work, by population.

Table 3 shows the distribution of tools. ChatGPT dominates in both populations, with secondary use of DeepL and Google Translate concentrated on language tasks. Gemini and Claude appear at lower frequencies, and at the end are the more specialised tools (Perplexity, Mistral) the use of which remains marginal in this study population.

Tool	Prépa ECG (%)	High School (%)
ChatGPT	42 (100.0%)	23 (100.0%)
Gemini	9 (21.4%)	8 (34.8%)
Claude	6 (14.3%)	7 (30.4%)
DeepL	13 (31.0%)	6 (26.1%)
Google Translate	5 (11.9%)	3 (13.0%)
Perplexity	9 (21.4%)	1 (4.3%)

Table 3. Tools used for academic work (multi-select; percentages exceed 100% by construction).

Table 4 reports the percentage of respondents using AI in each of the four main disciplinary clusters, alongside the Major Prépa (2025) reference distribution. The Fisher exact test is applied in two directions: between the prépa and high school populations of the present study, and between the prépa population and the Major Prépa national distribution.

Subject	HS	Prépa	MP 2025	HS vs Prépa (Fisher p)	Prépa vs MP (Fisher p)
Languages	78.3%	83.3%	78.2%	0.741 ns	0.556 ns
Eco/HGG/ESH or	56.5%	71.4%	67.3%	0.277 ns	0.730 ns

History-Geography					
Culture générale or Philosophy	65.2%	59.5%	57.3%	0.791 ns	0.871 ns
Mathematics	52.2%	69.0%	47.3%	0.192 ns	0.009 **

Table 4. Use of AI by academic discipline (percentage of respondents). Last two columns: Fisher exact test.

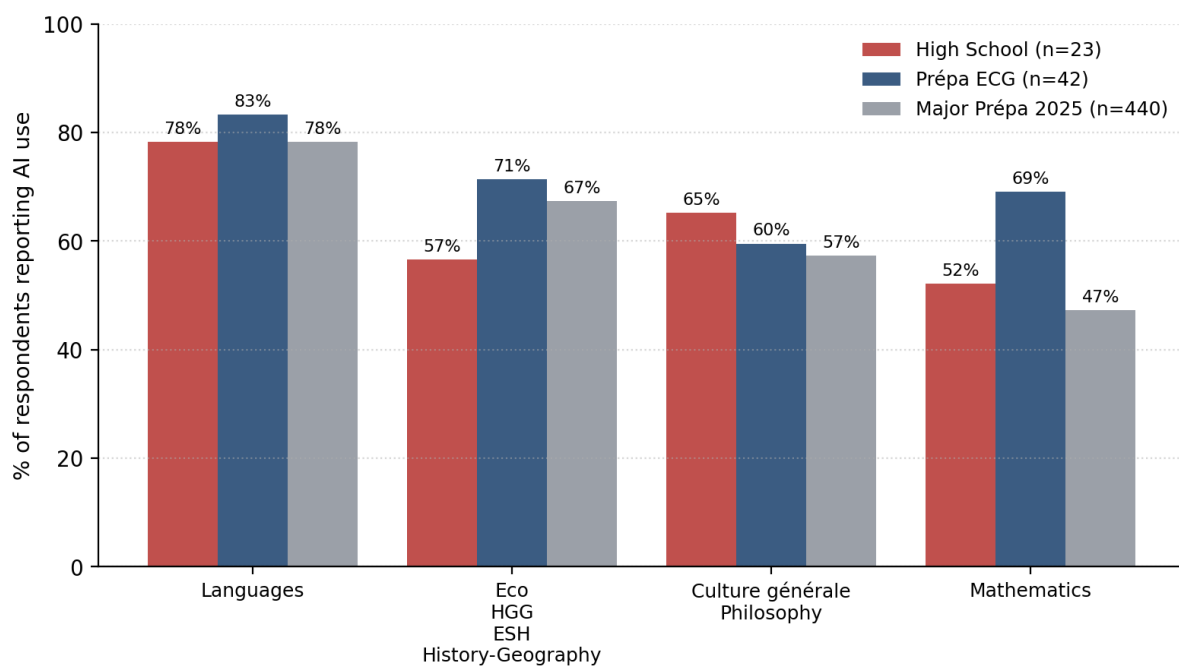


Figure 1. AI use across disciplines: comparison of high school, prépa ECG, and Major Prépa 2025.

The comparison between high school and prépa students within the present sample shows no significant differences across the four disciplines. The two populations report broadly similar disciplinary distributions of AI use. The comparison between the prépa sample and the Major Prépa national distribution reveals a significant divergence on Mathematics (prépa 69.0% vs MP 47.3%, $p = 0.009$).

The decomposition of AI use into three output-oriented behaviours (ideas generation, corrections and reformulations, full text generation) is reported in Table 5 and Figure 2. The threshold of Likert ≥ 5 is used to convert the present sample's Likert responses into binary

proportions, as described in section 2.4, in order to allow comparison with the yes/no format of the Major Prépa survey.

Output type	HS	Prépa	MP 2025	HS vs Prépa (Fisher p)	Prépa vs MP (Fisher p)
AI to generate ideas	60.9%	64.3%	55.2%	0.795 ns	0.329 ns
AI for corrections	65.2%	64.3%	43.9%	1.000 ns	0.014 *
AI to generate entire dissertations	30.4%	38.1%	46.4%	0.597 ns	0.334 ns

Table 5. Output-oriented use of AI (Q9a, Q9b, Q9c binarised at Likert ≥ 5). Fisher exact test.

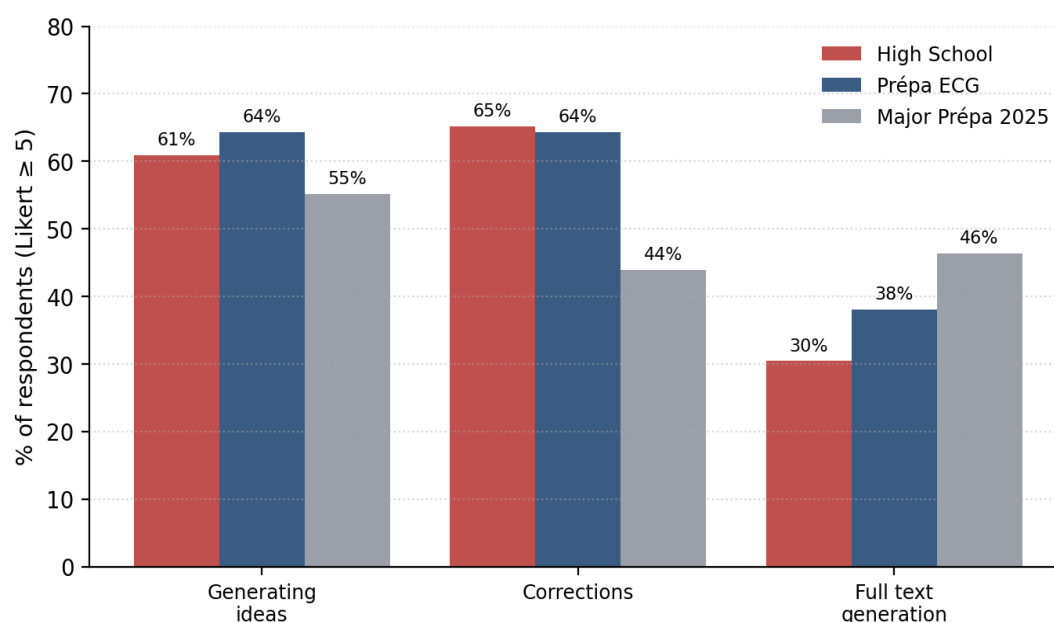


Figure 2. Decomposition of AI use by output type. Same threshold and comparison logic as Figure 1.

Within the present sample, the three output-oriented uses do not differ significantly between the high school and the prépa populations. The comparison with the Major Prépa distribution reveals a significant divergence on AI for corrections (prépa 64.3% vs MP 43.9%, $p = 0.014$). The pattern across the three behaviours is worth noting: ideas generation and corrections are common in all three samples, while full text generation, that is, the most substitutive form of use, is reported at lower rates in the present sample than in the Major Prépa one. This

direction of difference, although not always significant, is consistent across both the prépa and the high school groups and suggests a possible social desirability effect or socioeconomic and cultural differences between the families between the two surveys.

3.3 Cognitive effort and perceived autonomy

The third block of the questionnaire measures the dimensions of cognitive dependency: the moment at which the student turns to AI in solving a task, the metacognitive monitoring of the AI output, the perceived ability to perform the same task without the tool, the perceived effort invested in academic work. The Likert means and standard deviations for the two populations are reported in Table 6.

Item	Prépa mean (SD)	HS mean (SD)	Mann-Whitney p
Tries task alone before AI	5.69 (0.84)	3.57 (0.84)	<0.001 ***
Accepts AI output uncritically	2.64 (0.82)	4.48 (1.04)	<0.001 ***
Could work without AI	5.71 (0.92)	3.39 (1.03)	<0.001 ***
Detects AI errors	5.36 (0.93)	3.26 (1.01)	<0.001 ***
Less effort since using AI	3.64 (0.98)	4.39 (1.08)	0.013 *

Table 6. Cognitive effort and perceived autonomy: Likert means (SD) and Mann-Whitney U test.

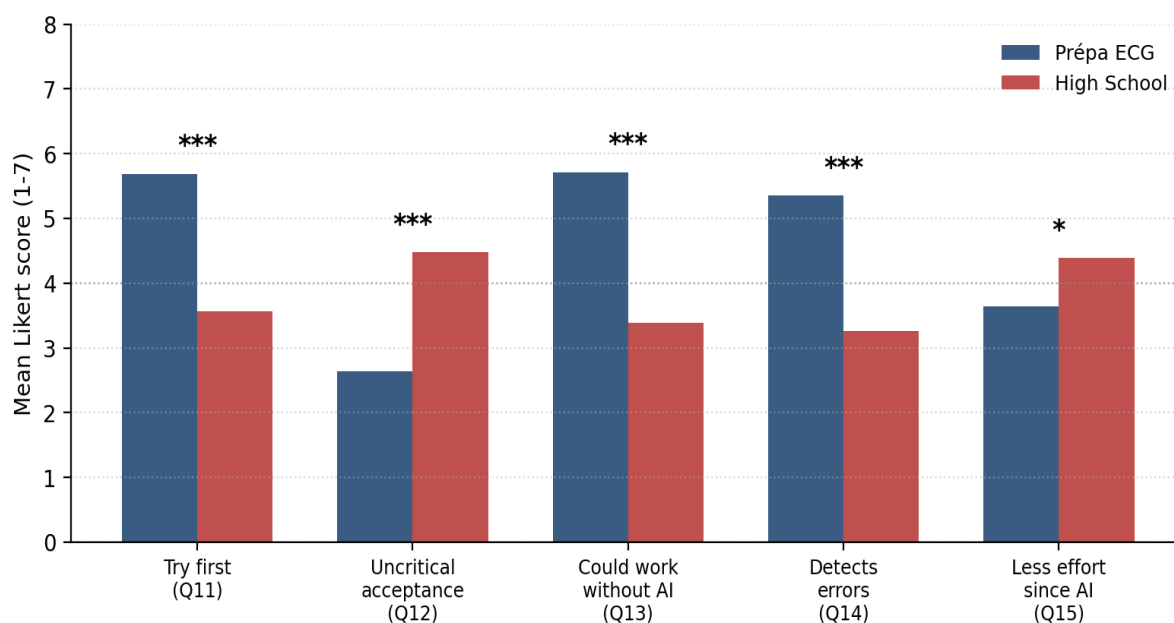


Figure 3. Cognitive dependency profile across the two populations. Significance markers: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

All five items in this block differ significantly between the two populations, all in the same direction: prépa students report higher autonomy on Q11 (mean 5.69 vs 3.57), lower uncritical acceptance of AI output on Q12 (mean 2.64 vs 4.48), higher perceived ability to work without the tool on Q13 (mean 5.71 vs 3.39), better error detection on Q14 (mean 5.36 vs 3.26), and a smaller perceived decline in effort since AI on Q15 (mean 3.64 vs 4.39). The direction of all five differences is consistent with the hypothesis that cognitively mature learners offload less of the germane load to the tool, monitor its output more critically, and preserve their autonomous capacity to a greater extent.

The qualitative material confirms and refines this pattern. In the in-depth exchanges with prépa students, several respondents explicitly describe a reflective sequence in which the tool is consulted only after a first autonomous attempt, with the AI output then compared against their own reasoning. In the high school material, the tutor observations document a different pattern, in which the tool is the first move rather than the last, and in which the AI output is accepted without independent verification. The convergence between the quantitative and the qualitative sources on this block is one of the most consistent findings of the analysis.

3.4 Cognitive maturity

The fourth block measures the four self-regulatory and metacognitive competences that, in

the framework of section 1.3, jointly define the cognitive maturity of the learner: planning before starting a task, monitoring progression during the task, evaluating the outcome after the task, and the ability to reason autonomously on unfamiliar material. Table 7 reports the means and standard deviations for each item. The four items being conceived as measures of the same underlying construct, a composite cognitive maturity index was computed as the within-respondent mean of Q16 to Q19.

Item	Prépa mean (SD)	HS mean (SD)	Mann-Whitney p
Plans approach	5.62 (1.10)	3.61 (1.08)	<0.001 ***
Monitors progression	5.67 (0.79)	3.70 (0.97)	<0.001 ***
Self-evaluates work	5.43 (0.97)	3.70 (0.97)	<0.001 ***
Autonomous reasoning	5.57 (1.11)	3.78 (1.00)	<0.001 ***
Composite maturity index (mean of Q16-Q19)	5.57 (0.45)	3.70 (0.58)	<0.001 ***

Table 7. Cognitive maturity: Likert means (SD), Mann-Whitney U test, and composite maturity index.

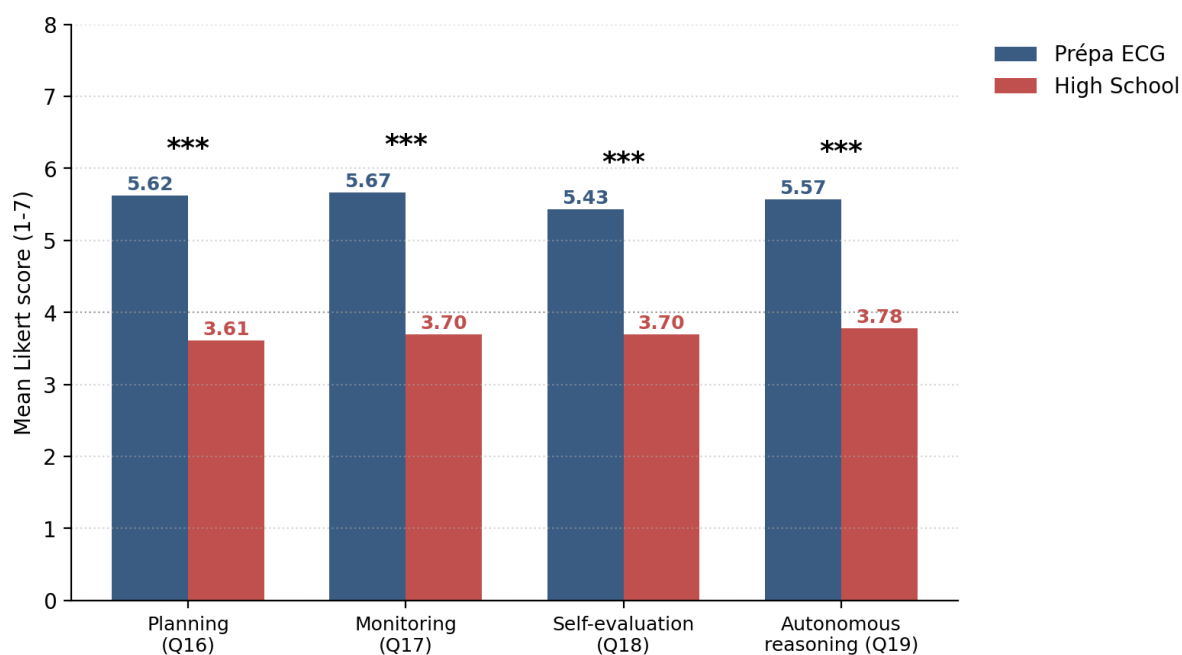


Figure 4. Cognitive maturity profile across the two populations.

All four items differ significantly between the two populations, and the composite maturity index produces a strong and highly significant gap between the prépa population (mean 5.57, SD 0.45) and the high school population (mean 3.70, SD 0.58), $p < 0.001$ (***). The two populations are exposed to the same technology in the same country at the same moment, but they differ structurally on the variable that the conceptual framework identifies as the modulating one. The size of the gap, more than one and a half points on a seven-point scale, suggests that the two populations occupy distinct regions of the maturity continuum and validates the choice of the comparative terrain made in section 2.1.

A correlation analysis at the individual level reinforces this reading. The Spearman correlation between the composite maturity index and item Q15 (perceived decline in effort since AI use) is negative and significant ($\rho = -0.33$, $p = 0.006$ **), indicating that more mature learners, irrespective of their population, report a smaller erosion of cognitive effort. This individual-level relationship is what makes the population-level comparison interpretable in causal terms within the limits of a cross-sectional design: the gap between the two populations on AI-related cognitive autonomy is not a population effect in the demographic sense, it is a maturity effect that travels with the variable across individuals.

3.5 Self-reported outcomes and attitudes

The fifth block measures the perceived effects of AI on academic performance and on durable understanding, the attitudinal positioning of the respondent toward the tool, and the projection of its long-term importance. The central distinction of this block, formulated in section 2.2.2, is between the perception that AI improves grades (Q20) and the perception that AI helps the student genuinely understand the material in a way that persists over time (Q21). Table 8 reports the descriptive statistics.

Item	Prépa mean (SD)	HS mean (SD)	Mann-Whitney p
AI improves grades	4.50 (1.04)	4.13 (0.87)	0.162 ns
AI helps durable understanding	4.05 (0.91)	3.57 (0.90)	0.065 †

AI is an ally	4.60 (0.96)	5.48 (0.99)	0.002 **
Worries about dependency	5.12 (1.29)	3.65 (0.88)	<0.001 ***
Worries about reasoning erosion	4.93 (1.33)	3.70 (1.02)	<0.001 ***
AI matters professionally	5.90 (0.82)	5.91 (0.90)	1.000 ns

Table 8. Outcomes and attitudes: Likert means (SD) and Mann-Whitney U test.

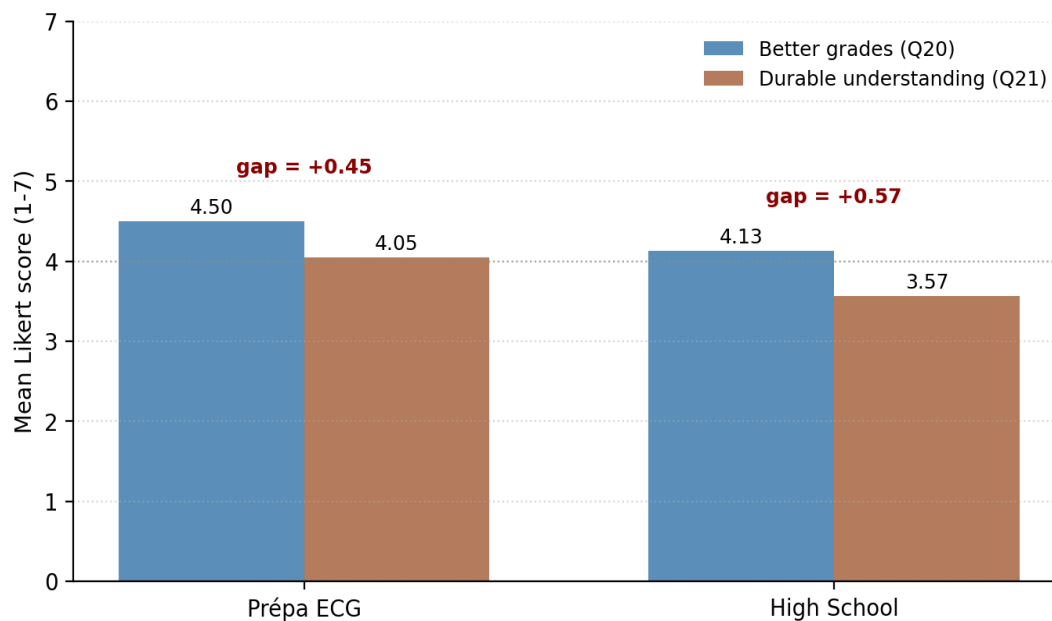


Figure 5. Perceived effects of AI on grades (Q20) versus on durable understanding (Q21), by population.

The divergence between Q20 and Q21 is positive in both populations: the perceived effect of AI on grades is consistently higher than the perceived effect on durable understanding. In the prépa population, the within-respondent gap (Q20 – Q21) has a mean of 0.45 points, with the gap being positive in 52% of respondents. In the high school population, the mean gap is 0.57 points, positive in 61% of respondents. The interpretation of this finding requires care. The gap is not, in itself, evidence of cognitive dependency. A student may very well believe, accurately, that AI improves a specific dimension of academic performance more than it improves the broader cognitive integration of the material. But the magnitude and the systematicity of the gap, observed in a majority of respondents in both populations, signals that the perception of a divergence between performance and learning is already widespread

among the students themselves. This is the empirical signature of the central paradox of this thesis, expressed by the respondents in their own self-assessment.

The two attitudinal items on long-term worry (Q23 on dependency and Q24 on reasoning erosion) produce significantly different distributions across the two populations. Prépa students report higher levels of concern on both items (Q23 mean 5.12 vs 3.65, Q24 mean 4.93 vs 3.70). This pattern is consistent with the framework of section 1.3: a more cognitively mature learner is also more capable of metacognitive distance, that is, of recognising the long-term cognitive risk of the tool that they use daily. The high school population, by contrast, shows lower mean scores on both items, which suggests not an absence of risk but a lower awareness of it.

The professional projection item (Q25) is the only one of the block on which the two populations do not significantly differ. Both groups report high agreement with the statement that mastering AI tools will matter for their professional future (prépa mean 5.90, HS mean 5.91). This convergence indicates that, regardless of cognitive maturity, students of both stages share a similar anticipation of the professional relevance of the technology, which is a useful baseline for the discussion of the managerial recommendations in chapter 4.

3.6 Summary of findings

The results of the quantitative analysis converge on three findings that the next chapter will discuss in light of the conceptual framework. First, on the descriptive items aligned with the Major Prépa survey, the present sample is comparable with the national reference distribution within the limits of statistical significance, which supports the external validity of the prépa sample. Second, the two populations of the present study do not differ significantly on the descriptive frequency and disciplinary distribution of AI use, but they differ markedly and significantly on the dimensions of cognitive effort, cognitive autonomy, metacognitive monitoring and cognitive maturity. The behavioural surface is therefore similar across the two populations, but the cognitive depth at which the tool is integrated differs substantially. Third, the perceived divergence between the effect of AI on grades and its effect on durable understanding is widespread in both populations, with a majority of respondents reporting a positive Q20 – Q21 gap.

Chapter 4. Discussion, Conclusions and Managerial implications

This thesis investigates the conditions under which generative AI in education supports deep learning rather than cognitive dependency. The question is examined through a comparative survey of 42 students of classes préparatoires économiques et commerciales and 23 high school students, two populations exposed to the same technology at the same moment but standing at different stages of cognitive development. The two groups use AI at similar frequency, in similar disciplines and for similar tasks. They differ significantly, however, on the cognitive operations that surround the tool. Prépa students more frequently attempt to solve the problem before consulting AI, check the output critically and retain the ability to work without it. High school students in the present sample more frequently consult AI first and accept its output with less verification. In both groups, a majority of respondents report that AI raises their grades more than it deepens their understanding. These results are consistent with the broader hypothesis that the cognitive effect of AI is determined by the maturity of the learner rather than by the technology itself. The crisis that has emerged in education, caused by the rapid adoption of generative AI outside any institutional framework, therefore calls for a redefinition of homework and a redesign of the evaluation system: a grade that rewards a correct answer cannot tell whether the student understood the material or simply prompted an algorithm. The thesis argues for three converging responses: the deliberate integration of generative AI into the curriculum, the early training of students in its critical and reflective use, and the redesign of evaluation around the cognitive operations that genuinely produce deep learning.

4.1 Generative AI as a daily tool in constant evolution

The present survey conducted in the first months of 2026 shows a massive use of generative AI among both prépa and high school students. It aligns with the one of Major Prépa from 2025 and other surveys in France, showing that more than 80% of high school students use ChatGPT or equivalent technology for their homework. The Born AI 2025 barometer found that 93% of 18-25 year olds have used a generative AI tool in the previous six months in France. In Italy, 65% of 16-18 year olds in 2025 use it to study (Skuela.net / NoPlagio.it 2025) and in the United Kingdom, 92% of undergraduate students in 2025 use it for their academic work

(HEPI 2025). Even a 2024 survey conducted in Nouvelle-Aquitaine found that more than 90% of seconde students had already used a generative AI tool for their homework, showing that this extremely high usage of generative AI is not a new phenomenon but appeared very soon after the appearance of ChatGPT on the market.

The use of generative AI for full text generation of a given assignment is reported at lower rates in the present sample than in the Major Prépa one, although the difference does not reach statistical significance. This suggests either a social desirability effect specific to a population in contact with a tutoring structure, or a genuine difference reflecting a better understanding of the limits of the tools by students who have had access to them and used them over the past three to four years, consistent with the qualitative material gathered.

For the majority of the usage categories tested, the results obtained here in 2026 for both studied populations, prépa ECG and high school students, did not differ significantly between each other and from the prépa students from the Major Prépa from 2025. Two differences were notable. The first one concerns the use of AI in mathematics, which is significantly more frequent in the present sample than in the Major Prépa one (69.0% vs 47.3%, $p = 0.009$). The second concerns the use of AI for corrections and reformulations of one's own writing, which is also more frequent in the present sample (64.3% vs 43.9%, $p = 0.014$). Even if the average results were very similar between the prépa and high school students in this study, only the prépa results were significantly different from the Major Prépa survey, likely due to the small sample size of the high school group and the lack of statistical power. Therefore, one explanation for the observed differences could be the small sample size and the differences in the recruitment channels. Nevertheless, it is likely that the differences are genuine and that they could be explained with the improvement of the quality and the rapid technical evolution of the tools themselves between the moment of the Major Prépa survey, performed before May 2025, and the moment of the present one performed between February and March 2026. The Major Prépa data was collected at a stage when ChatGPT, the dominant model used by all students, was still running on the GPT-4 family, with documented limitations on mathematical reasoning and a non-negligible rate of hallucinations on technical content. Since then, the release of GPT-5 in August 2025 (after the Major Prépa survey) and its successive iterations, GPT-5.1 in November 2025, GPT-5.2 in February 2026 and GPT-5.4 in March 2026, have produced improvement in mathematical performance, with reported scores of 94.6% on

the AIME 2025 benchmark, and a marked reduction in hallucination rates. Competing tools such as the newly introduced and very powerful tool Claude (Opus 4.7 model) have followed the same trajectory. The same period has also seen the diffusion of fine-grained reformulation and rewriting features, which make AI substantially more useful for the iterative correction of one's own academic writing than it was one year earlier. The higher rates observed in the present sample on mathematics and on corrections and reformulations may therefore reflect the fact that the tools have, in the interval, become reliable enough on these two specific tasks to justify a more frequent recourse by students who report in the qualitative material that they were previously disappointed by them.

From this first analysis it is evident that the question is no longer whether students use AI, since they all do, but how the tool is being integrated into their cognitive routines and what this integration produces over time.

4.2 Behavioural convergence but cognitive divergence between the two populations

The most striking pattern in the quantitative results is the contrast between block 2 of the questionnaire, which measures how students use AI, including frequency, tools and disciplines, and blocks 3 and 4, which measure how deeply the tool is integrated into their cognitive work.

On the second block, the two populations do not differ significantly on frequency of use, disciplinary distribution or output-oriented behaviours. From the outside, a *prépa* student and a high school student look like users of the same tool, in the same disciplines, at roughly the same intensity. On the third and fourth blocks, by contrast, every single item differs significantly between the two populations, and the direction of the difference is systematically the same: *prépa* students more frequently attempt the task autonomously before consulting AI (mean 5.69 vs 3.57), accept the AI output less uncritically (mean 2.64 vs 4.48), preserve their perceived capacity to work without the tool (mean 5.71 vs 3.39), detect AI errors more reliably (mean 5.36 vs 3.26), and report a smaller decline in cognitive effort since they started using AI (mean 3.64 vs 4.39). The composite cognitive maturity index produces a gap of more than one and a half points on a seven-point scale between the two populations.

This pattern corresponds to the working hypothesis of this study. The frequency of use, taken

in isolation, is not the cognitively relevant variable. Two students may interact with ChatGPT five times a day and yet be doing entirely different things with the tool, depending on whether they offload extraneous load (which is beneficial) or germane load (which is harmful), and depending on whether the offloading is supervised by metacognitive monitoring or executed by default. The behavioural surface, which is what surveys typically measure and what the public debate fixates on, often hides the cognitive depth at which the tool is actually integrated. The present results document this hidden depth: at equal frequency and equal disciplinary distribution, the two populations stand at different positions on the maturity continuum, and this difference is what determines whether AI use produces a reflective or a substitutive cognitive pattern.

This finding sits within a broader empirical pattern that has begun to emerge in the literature on AI use across age groups. Sublime and Renna, in a preprint from 2024 (the author of this thesis did not find the paper published in a peer-reviewed journal), conducted a survey of 395 students aged 13 to 25 in France and Italy divided into four age groups. They found that the proportion of students who systematically rework AI-generated answers rises from 20% in the 13-16 age group to nearly 50% in older groups ($p = 0.0017$), and that older students show greater frequency and quality of AI use across both humanities and scientific topics. Gerlich (2025), in a study of 666 participants, reports that younger users (aged 17-25) exhibit higher cognitive offloading and lower critical thinking scores than older users. Education played a protective role: participants with higher levels of education preserved their critical thinking skills even when they used AI frequently. The direction of the present findings corroborates these two precedents on a French-only sample.

What the present study adds, beyond replication, is the dissociation between age and selection. The two populations of this thesis overlap in age range with the upper end of Sublime and Renna's spectrum and with the lower end of Gerlich's, but they are separated by a selection mechanism that concentrates self-regulatory and metacognitive competences in the *prépa* group, while the high school students are still in construction of their cognitive capacity. The observed effect can therefore be attributed to maturity itself, rather than to age, which merely tends to track maturity. The correlation analysis at the individual level reinforces this interpretation. The Spearman correlation between the composite maturity index and the perceived decline in effort since AI use (Q15) is negative and significant across the pooled

sample ($p = -0.33$, $p = 0.006$). More mature learners, irrespective of whether they belong to the prépa or the high school population, report a smaller erosion of cognitive effort. The maturity variable therefore travels with the individual, not with the institutional label. This is what makes the comparison between the two populations interpretable as a maturity effect rather than as a difference between age groups or school types: the gap is not produced by the prépa label itself, but by the cognitive characteristics that the prépa selection process concentrates and that high school students are, on average, still in the process of building.

The qualitative material supports the quantitative pattern. The in-depth exchanges with prépa students document a reflective sequence in which the tool is consulted after a first autonomous attempt and in which the AI output is compared with the student's own reasoning before being integrated. Several respondents describe explicitly the moment at which they refuse to let AI generate the answer, because they recognise the cognitive cost of skipping the autonomous attempt. The tutor observations and the answers of the high school population document a different pattern, in which the tool is often the first move rather than the last and in which the AI output can be accepted without independent verification. This is the difference between offloading the germane load with metacognitive supervision and offloading it without. The convergence between the quantitative and the qualitative material on this block therefore supports the quantitative data conclusions.

4.3 The perceived gap between grades and durable understanding

The fifth block of the questionnaire showed that in both populations, the perceived effect of AI on grades exceeds the perceived effect of AI on durable understanding. The within-respondent gap (Q20 – Q21) has a positive mean in both groups (0.45 in the prépa population, 0.57 in the high school population) and is positive in a majority of respondents in both groups (52% in the prépa population, 61% in the high school population). The two populations therefore agree on the existence of the gap, even though they disagree on almost every cognitive dimension that surrounds it.

This finding can be defined as an empirical signature of the AI paradox: AI improves short-term performance while removing the cognitive effort that produces long-term learning. The meta-analytic evidence of Deng and colleagues (2025) already documents similar divergence, a positive average effect on academic achievement combined with a measurable reduction of the mental effort invested in the task. A majority of respondents in both populations in the

current study perceive that AI lifts their grades more than it lifts their durable understanding. This finding should not be interpreted as proof of cognitive damage, because a self-reported gap between two perceived effects is small and not a direct measurement of such an effect. But the systematicity of the gap, observed in a majority of respondents in both populations and reported despite the obvious social desirability cost of admitting such a gap on one's own learning, gives it weight as a relevant finding.

The two attitudinal items on long-term worry (Q23 on dependency and Q24 on reasoning erosion) support it. Prépa students report significantly higher levels of concern on both items (mean 5.12 vs 3.65 on dependency, mean 4.93 vs 3.70 on reasoning erosion). These findings are consistent with the notion that more cognitively mature learners are also more capable of metacognitive distance, that is, of recognising the long-term cognitive risk of the tool that they use daily. The high school population, in contrast, reports lower levels of concern on both items, which does not mean an absence of risk but a lower awareness of it. This asymmetry is itself a finding of pedagogical relevance: the population that is the most exposed to the risk, by virtue of being still in the process of building the cognitive foundations that protect against it, is also the population that reports the lowest awareness of the risk.

The professional projection item (Q25), the only one of the block on which the two populations converge, shows that the future-oriented narrative on the importance of mastering AI is shared across the two stages of the curriculum. The convergence on professional relevance combined with the divergence on cognitive worry is important for the managerial discussion: students of both stages agree that they will need to master AI for their future, but mostly the prépa students display the metacognitive distance that allows mastery rather than substitution.

4.4 The crisis of homework and students' knowledge and skills assessment: managerial recommendations

Overall, the findings here and the data in the literature invite a reframing of the public debate on AI in education. The debate often focuses on the behaviour of the students: do they use the tool too much, do they use it for the wrong tasks, do they cheat with it. The present results show that the two populations of the study use the tool at comparable intensity, in comparable disciplines, for comparable output-oriented tasks. Therefore, the two populations differ not on the quantity of their AI use but on the quality of the cognitive operations that

surround it, and this quality is itself a function of the maturity already built. If the cognitive effect of AI depends on the maturity of the learner and not on the technology itself, then no policy that targets the technology alone can resolve the cognitive question. Banning the tool or tolerating its unsupervised use will not build maturity.

Nguyen and colleagues (2025), in a policy document analysis published in *Information and Learning Sciences*, examine guidelines from major intergovernmental organisations (OECD, UNESCO, European Commission, Council of Europe) to identify a global consensus on how learning design and assessment should evolve in a world where generative AI is embedded in both education and work. The authors find broad agreement that learning outcomes must shift toward higher-order thinking skills (critical thinking, collaboration, transfer of learning) and AI literacy, and that assessment must be rethought around four priorities: cheat-proof formats, formative evaluation, integration of AI into assignments and pedagogical practices, and ethical safeguards on data, equity and transparency. Their core argument is that the integration of generative AI calls for hybrid intelligence approaches where AI augments rather than replaces human cognition, and that current policies still lack concrete guidance on how to scaffold transferable cognitive skills, leaving educators and institutions to implement these principles with limited operational support. This paper, though, does not propose concrete solutions for the identified gap. Based on the results of this thesis, the literature and the own pedagogic experience of the author, the following managerial implications can be formulated.

The **first managerial implication** of this work concerns the place of homework in the learning process. If the cognitive operations that produce deep learning can be offloaded to AI in a few seconds, then the asynchronous written assignment, which has been the cornerstone of the academic curriculum for two centuries, loses its function both as a training device and as an assessment instrument. This conclusion is shared by a growing body of literature: Cotton and Shipway (2024) frame the question as an integrity crisis that institutions must address through new evaluative formats; Sublime and Renna (2024) recommend rethinking homework explicitly in light of the fact that AI assistance has become near-universal; and Mollick and Mollick (2023) argue that the integration of AI in education must be accompanied by a shift away from take-home assessment and toward authentic, interactive forms of evaluation.

The pedagogical experience of the author converges on the same conclusion. In the tutoring of prépa ECG and high school students by the author or the tutors of his company Prépa

Rationnelle, no homework is given between sessions. Exercises are solved on paper, by hand, in real time during the session, which reproduces the cognitive conditions of the written concours and makes AI assistance physically unavailable. For prépa students and for first-year students of OTERIA Cyber School, where the author teaches mathematics and Python, he developed a dedicated platform, <https://www.preparationnelle.com/>, on which exercises are completed in real time under supervision during the tutoring or practical lessons, within a limited and bounded period, which removes most of the opportunities for substitutive AI use. These pedagogical choices are not a definitive answer to the question raised by AI in education, but they illustrate a principle that the present results support: the cognitive operations that produce learning must take place under conditions where the tool cannot replace them, which in practice means a return to synchronous, supervised and time-bounded forms of academic work.

The **second managerial implication** is that the curriculum must integrate explicit training on how to use AI well. The premise of this recommendation is that the question of whether students will use the tool is settled: they do, in proportions approaching 100%, as the Born AI 2025 barometer, the HEPI 2025 survey and the present data converge to show. The relevant question is therefore not whether to accept the use of AI but how to teach its productive use. The literature on this point is consistent. Kasneci and colleagues (2023) argue that large language models can become powerful educational allies if their integration is intentional and pedagogically informed. Mollick and Mollick (2023) develop a series of concrete strategies through which AI can operate as a force multiplier for effective teaching, provided that students are explicitly trained in the reflective rather than substitutive mode of use. Sublime and Renna (2024) reach the same conclusion: the absence of explicit training in the critical evaluation of AI outputs is one of the main drivers of the cognitive gap they observe between younger and older students. The results of the present thesis have led the author to prepare for the next academic year for all students of Prépa Rationnelle a dedicated webinar, recorded and available on demand, presenting the findings of this study and articulating the good practices of AI use for deep learning. The webinar will distinguish explicitly the offloading of extraneous load, which is beneficial and to be encouraged, from the offloading of germane load, which erodes the cognitive operations that produce durable learning. The intention is to give students the metacognitive vocabulary to recognise, in their own daily practice, the

moment at which the tool stops supporting learning and starts replacing it.

The **third managerial implication** concerns the assessment system itself. An assessment system that rewards the production of a correct answer cannot distinguish a student who has understood from a student who has prompted an algorithm, and it therefore cannot reward the cognitive operations that produce durable learning. In a UNESCO opinion piece published in May 2025, Desai argues that the rise of generative AI tools has triggered an assessment crisis in education and that the response cannot be a technological arms race between AI detection and AI generation, which is already a lost battle, but a fundamental redesign of what assessment measures. The argument is that traditional evaluations rely on outputs (essays, problem sets) that AI now produces in seconds, which means these outputs no longer reliably indicate individual understanding or effort. Desai suggests that this crisis is also an opportunity, because conventional assessment has long been criticised for rewarding memorisation and formulaic responses rather than deep learning, and AI has made the limits of these methods undeniable. The proposed redesign consists of four directions: assessing higher-order skills that AI cannot fully replicate, namely critical thinking, creative problem-solving and ethical reasoning; shifting the focus from final artifacts to the learning process through progress reports, iterative submissions, peer reviews and metacognitive reflection; introducing dialogue-based and oral defence formats (thesis defences, Socratic seminars, structured interviews) where intellectual ownership must be demonstrated in real time; and anchoring assignments in students' personal contexts and local communities, which AI cannot easily substitute for.

In the current teaching practice of the author, mathematics and Python evaluations are conducted in real time, on paper and by hand, in conditions that reproduce the cognitive constraints of the written concours and that make substitutive AI use impossible. This form of evaluation rewards the autonomous mastery of the discipline and remains indispensable, because the foundations of mathematical and algorithmic reasoning cannot be outsourced without being lost. Based on the Desai suggestion and the outcome of the survey, a key competence to be mastered by the students is the capacity to use AI productively in the resolution of complex problems, which will be a baseline requirement of nearly every profession these students enter. The assessment system must therefore evolve toward a dual format: alongside the traditional unassisted examination, the author plans to introduce,

particularly for the OTERIA Cyber School students, a second form of evaluation explicitly designed to test the quality of AI use. The format is a supervised, time-bounded session during which students are given a complex problem (a non-trivial mathematical modelling task or an open-ended Python project) and are allowed, even required, to use generative AI. The evaluation does not grade the final answer in isolation, because the answer is now trivially available. It grades the reasoning trajectory, namely the framing of the problem, the formulation of the prompts, the critical evaluation of the AI output, the identification of its errors, the reformulation in the student's own terms, and the integration of the result into a coherent solution that the students must present orally and answer questions on. This format directly operationalises the three proxies of deep learning: transfer to a new situation, re-explanation in one's own words, connection to existing knowledge, and applies them to the AI-augmented setting that will dominate the professional life of the students. The combination of the two formats, the unassisted exam that protects the foundations and, at least for now, reproduces the concours for the Business Schools, and the AI-assisted preparation of an oral exam that tests the productive integration of the AI tool, gives the curriculum the instruments to develop both competences and to evaluate them rigorously.

Bibliography

Akgun, S. & C. Greenhow (2022), "Artificial intelligence in education: addressing ethical challenges in K-12 settings", *AI and Ethics*, vol. 2, no. 3, p. 431-440.

Biggs, J. B. (1987), *Student Approaches to Learning and Studying*, Australian Council for Educational Research, Melbourne.

Born AI (2025), *Baromètre Born AI 2025: usages des outils d'intelligence artificielle générative chez les 18-25 ans en France*, Agence Heaven, Paris, June.

Cotton, D. R. E., P. A. Cotton & J. R. Shipway (2024), "Chatting and cheating: ensuring academic integrity in the era of ChatGPT", *Innovations in Education and Teaching International*, vol. 61, no. 2, p. 228-239.

Danyaro, K. U., M. S. Liew & J. Jaafar (2025), "Hallucinations in large language models:

implications for educational settings”, *International Journal of Teaching, Learning and Education*, 4(6), 13–19. <https://doi.org/10.22161/ijtle.4.6.2>

Deng, R., M. Jiang, X. Yu, Y. Lu & S. Liu (2025), “Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies”, *Computers and Education*, vol. 227, art. 105224.

Desai, H. (2025), “What’s worth measuring? The future of assessment in the AI age”, UNESCO IdeasLAB, Paris, 19 May, <https://www.unesco.org/en/articles/whats-worth-measuring-future-assessment-ai-age>.

Entwistle, N. J. (1981), *Styles of Learning and Teaching: An Integrated Outline of Educational Psychology*, John Wiley & Sons, Chichester.

Flavell, J. H. (1979), “Metacognition and cognitive monitoring: a new area of cognitive developmental inquiry”, *American Psychologist*, vol. 34, no. 10, October, p. 906-911.

Gerlich, M. (2025), “AI tools in society: impacts on cognitive offloading and the future of critical thinking”, *Societies*, vol. 15, no. 1, art. 6.

HEPI (2025), *Student Generative AI Survey 2025*, Higher Education Policy Institute, Oxford.

Ji, Z., N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto & P. Fung (2023), “Survey of hallucination in natural language generation”, *ACM Computing Surveys*, vol. 55, no. 12, art. 248, March, p. 1-38.

Jose, B., J. Cherian, A. M. Verghis et al. (2025), “The cognitive paradox of AI in education: between enhancement and erosion”, *Frontiers in Education*. <https://doi.org/10.3389/fpsyg.2025.1550621>

Kasneci, E., K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günemann, E. Hüllermeier et al. (2023), “ChatGPT for good? On opportunities and challenges of large language models for education”, *Learning and Individual Differences*, vol. 103, art. 102274.

Lancaster, T. (cited in *The Guardian*, 2025), “Revealed: thousands of UK university students caught cheating using AI”, *The Guardian*, 15 June 2025, reporting figures of 5.1 confirmed cases of AI cheating per 1,000 students in 2023-24 (up from 1.6 the previous year), based on a survey of 131 UK universities.

Liu, Z., H.Zuo, and Y.Lu. 2025. "The Impact of ChatGPT on Students' Academic Achievement: A Meta-Analysis." *Journal of Computer Assisted Learning* 41, no. 4: e70096. <https://doi.org/10.1111/jcal.70096>.

Major Prépa (2025), Enquête sur l'utilisation de l'intelligence artificielle par les étudiants de classes préparatoires, May, <https://major-prepa.com/actualite/80-des-etudiants-de-prepa-utilisent-lia-lenquete-de-major-prepa-devoile-un-usage-massif-mais-nuance/>.

Marton, F. & R. Säljö (1976), "On qualitative differences in learning: I – outcome and process", *British Journal of Educational Psychology*, vol. 46, no. 1, February, p. 4-11.

Mollick, E. & L. Mollick (2023), "Assigning AI: seven approaches for students, with prompts", The Wharton School Research Paper, University of Pennsylvania, Philadelphia, SSRN abstract 4475995 / arXiv:2306.10052, 23 September.

Nguyen, A., A. T. Duong, D. T. B. Nguyen, V. T. T. Lai & B. Dang (2025), "Guidelines for learning design and assessment for generative artificial intelligence-integrated education: a unified view", *Information and Learning Sciences*, DOI 10.1108/ILS-11-2024-0148.

Nouvelle-Aquitaine Region (2024), survey on the use of generative AI by students of seconde in the region, cited by Public Senat:

<https://www.publicsenat.fr/actualites/education/intelligence-artificielle-un-rapport-du-senat-recommande-un-cadre-dusage-pour-mettre-lia-au-service-de-lecole>

Piaget, J. (1964), "Part I: cognitive development in children: Piaget development and learning", *Journal of Research in Science Teaching*, vol. 2, no. 3, p. 176-186.

Portex, M., C. de la Higuera & J.-J. Vie (2025), *Baromètre 2025: enseignants et intelligence artificielle*, Ecolhuma, Paris, May.

Risko, E. F. & S. J. Gilbert (2016), "Cognitive offloading", *Trends in Cognitive Sciences*, vol. 20, no. 9, September, p. 676-688.

Skuola.net & NoPlagio.it (2025), *Indagine sull'uso dell'intelligenza artificiale generativa tra gli studenti italiani delle scuole superiori*, Skuola.net, Rome.

Sublime, J. & I. Renna (2024), "Is ChatGPT massively used by students nowadays? A survey on the use of large language models such as ChatGPT in educational settings", arXiv preprint, arXiv:2412.17486, December (preprint, not peer-reviewed).

Sweller, J. (1988), "Cognitive load during problem solving: effects on learning", *Cognitive Science*, vol. 12, no. 2, April-June, p. 257-285.

Vygotsky, L. S. (1978), *Mind in Society: The Development of Higher Psychological Processes*, M. Cole, V. John-Steiner, S. Scribner & E. Souberman (éd.), Harvard University Press, Cambridge, MA.

Wood, D., J. S. Bruner & G. Ross (1976), "The role of tutoring in problem solving", *Journal of Child Psychology and Psychiatry*, vol. 17, no. 2, April, p. 89-100.

Zawacki-Richter, O., V. I. Marín, M. Bond & F. Gouverneur (2019), "Systematic review of research on artificial intelligence applications in higher education: where are the educators?", *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, art. 39.

Zimmerman, B. J. (2002), "Becoming a self-regulated learner: an overview", *Theory into Practice*, vol. 41, no. 2, p. 64-70.