SQUARE 2022 - 2025

Impact report

Design : ©Julia Fraud

With a large-scale sample of over 750 middle schoolers and 630 young adults from disadvantaged backgrounds in Ile-de-France, this report offers unique insights into gaps in media literacy, critical thinking, and civic confidence, while also assessing the impact of Square's political and media literacy workshops.

SQUARE

Education workshops for young people in disadvantaged neighborhoods in Ile-de-France

This report has been authored by **Dr. Melisa Basol**, a social psychologist specializing in the development, testing, and scaling of efforts to combat misinformation. **Dr. Basol**'s extensive research and practical applications in this field are aimed at strengthening Square's understanding and mitigation of the impact of false information through innovative strategies and evidence-based interventions.

This report includes contributions from **Dr. Manon Berriche** (Annex 1), postdoctoral researcher in sociology at Sciences Po medialab, and **Dr. Sacha Altay**, postdoctoral researcher in behavioural sciences at the University of Zurich.

The initial protocol and evaluation questionnaires were designed in collaboration with: **Ilf Bencheikh**, Director of Training at J-PAL Europe, **Dr. Simon Briole**, Lecturer and Researcher in Economics at the University of Montpellier, and **Dr. Quentin Daviot**, economist specializing in the evaluation of educational programs. The latter collected and analysed the data and produced findings for the 228 young adults involved in Square's workshops who responded the questionnaires in 2023 (Annex 2). They also participated in the first RCT evaluation carried out by Square on its 2021 pilot project "Digital resilience community ".

Sommaire

Executive summary

With a large-scale sample of over 750 middle schoolers and 630 young adults from disadvantaged backgrounds in Ile-de-France, this report offers unique insights into gaps in media literacy, critical thinking, and civic confidence, while also assessing the impact of Square's political and media literacy workshops.

Key findings

• Participants reported feeling more engaged in political discussions and more confident in expressing their opinions post-workshop. The strongest gains were in political self-expression, with more participants believing they had opinions worth listening to.
• Participants translated intention into action, maintaining their commitment to critical engagement with online content for up to three months.
• Participants demonstrated clear improvements in defining key media literacy concepts, with gains in understanding confirmation bias, and algorithmic filtering, sustained at follow-up.
• After the workshop, most of the participants reported checking sources more frequently, and seeking out diverse perspectives, suggesting an increased awareness of verification practices.
• The results suggest that misinformation susceptibility in this cohort is less about outright belief and more about uncertainty.

Implications and recommendations

The findings from this unprecedented data collection effort illustrate that short-term media literacy interventions can effectively raise awareness and enhance conceptual understanding critical to navigating digital environments. This report also highlights the importance of longitudinal assessment in measuring the lasting impact of such initiatives. While initial results suggest positive shifts, ongoing data collection and more follow-up analyses are crucial to determining whether these gains translate into sustained real-world application and long-term behavioural change.

To build on these gains, future interventions should:
• Expand civic engagement components to increase confidence in navigating political information.
• Sustain exposure over time, as one-off workshops alone may not be enough to drive long-term behavioural change.
• Integrate more hands-on misinformation detection exercises to strengthen real-world critical thinking skills.
• Leverage prebunking strategies to equip the audience with resilience against emerging misinformation narratives and techniques.

Part I: Understanding the gaps in media literacy

Between 2023 and 2025, Square partnered with local NGOs, schools and employment organisations to support middle schoolers and young adults from disadvantaged neighbourhoods in Clichy-sous-Bois, Les Mureaux and Mantes-la-Jolie, areas marked by high poverty and unemployment.

The total baseline sample reached up to 750 and 630 participants, though the exact number of answers varies across different metrics. Some items were introduced, removed, optimised, or adjusted throughout data collection, and participants could choose not to answer certain questions, leading to slight variations in sample sizes for specific measures. The number of responses will be specified below for each dimension "(N = x)".

This part examines participants' baseline abilities before taking part in Square's workshops.

The first section examines the baseline findings from middle schoolers, followed by insights from young adults in a second section. The findings reveal major gaps in critical thinking, credibility assessment, and misinformation detection, underscoring the urgent need for multi-layered and evidence-based initiatives. Specifically, the data highlights high distrust in institutions, endorsement of conspiracy beliefs, and motivated reasoning, where many reject information that contradicts their views. Combined with uncertainty in evaluating online content and weak verification habits, these challenges make a clear case for structured media literacy education. These baseline insights highlight key vulnerabilities and establish a foundation for assessing the long-term impact of our workshops, which will be examined in the second part.

Section 1: Media literacy gaps in middle schoolers

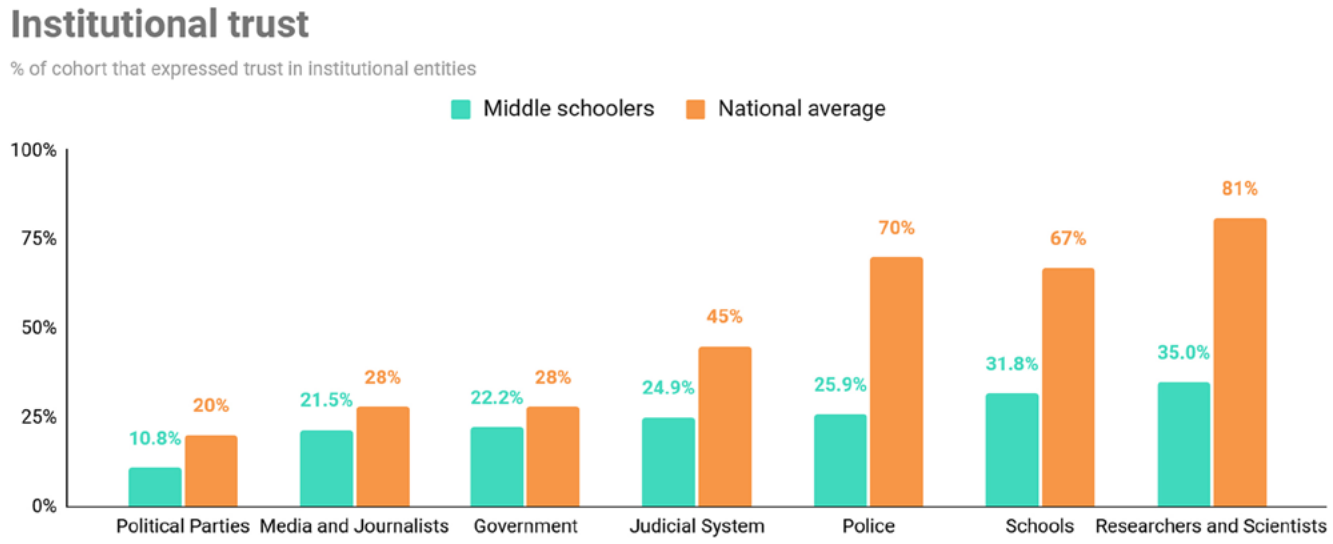
Sample insights

The sample comes from working with 2 middle schools in the Paris area: collège Jules Verne in Les Mureaux and collège de Gassicourt in Mantes-la-Jolie. The participants came from each class level (from 6ème to 3ème, or 11 to 15 years old). Both schools are REP+ where education is reinforced to fight against school dropout and failure. In terms of IPS (indicator of social positioning), Jules Verne averages 80.60 and is located far below the average (106,5). Gassicourt is slightly higher with 81.70. Both schools are ranked below 5000 out of 6615 in results of the final middle school degree (Brevet des collèges) with 83.96% success rate for Gassicourt and 77.86% for Jules Verne.

Institutional trust (N = 408)

There is a profound institutional trust gap among our cohort compared to the national average. A deep scepticism toward societal structures is evident, with their average trust being just 24.5%, barely half of the 48.4% national average[1] (all ages).

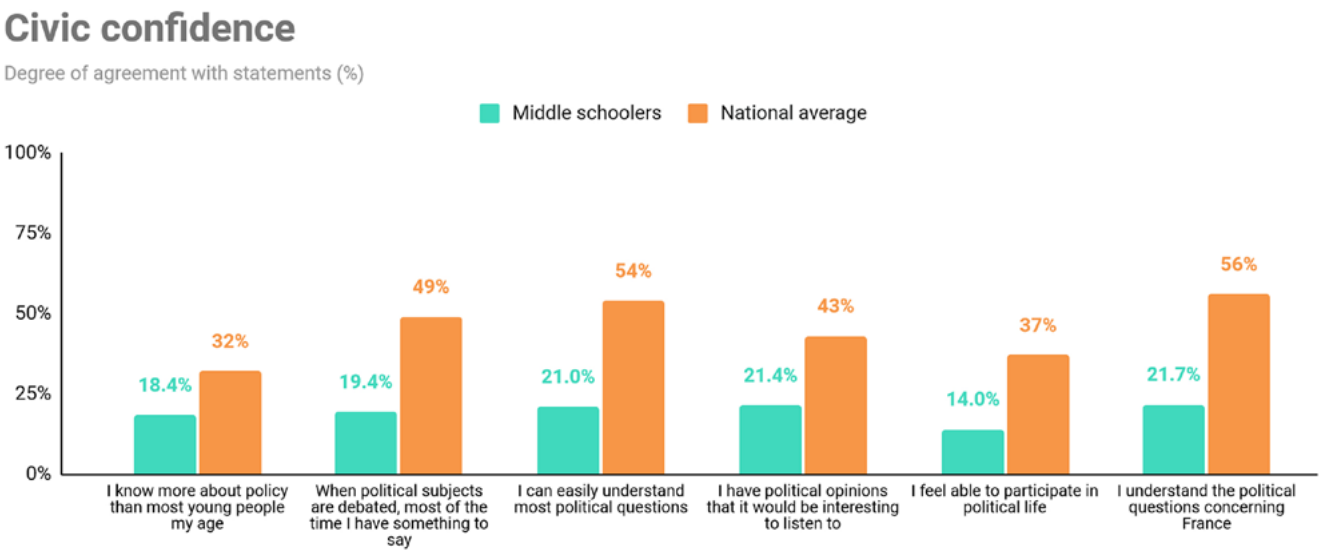
Critically, the least trusted institution is political parties, reflecting strong disillusionment with formal governance. The starkest disparities appear in law enforcement (25.9% vs. 70%) and the judicial system (24.9% vs. 45%), suggesting that our cohort may feel particularly alienated from the systems tasked with upholding justice and security. This data paints a troubling picture: institutions that are meant to provide stability and fairness are failing to earn the trust of our participants. While the most trusted messengers among middle schoolers are researchers and scientists (35%) and schools (31.8%), even these figures fall significantly below the national averages (81% and 67%, respectively).



Civic confidence (N = 392)

Middle schoolers display alarmingly low civic confidence[2], with trust in their political knowledge and influence falling far below the national average. Our research suggests that they feel less informed, less heard, and less capable of participating in political life compared to the broader population of French teenagers.[3] While many people nationally express confidence in understanding and discussing political issues, our cohort struggles to see themselves as knowledgeable or influential. The widest gaps appear in political comprehension and ability to participate, where national confidence is more than double that of our cohort.

Without intervention, these low confidence levels risk cementing long-term political disengagement or, in the worst cases, fostering radicalisation.[4] When young people feel excluded from formal political participation and unheard by institutions, some may seek alternative, sometimes extreme, avenues to express their frustration and reclaim a sense of agency. Empowering marginalised youth with civic knowledge, engagement opportunities, and a belief in their ability to effect change is critical.



Social media usage and trust (N = 556)

A noted paradox in our findings is the coexistence of high social media usage with low trust in the information presented on these platforms. 54% report low or no confidence in the information they see on social media.

Research suggests that social media plays a dual role in fulfilling basic human needs for connection while also being a conduit for misinformation. It reflects a nuanced relationship between technology and social behaviour, stating that "the primary benefit is social connection, and that's true for teens who are connecting with friends they already have or making new connections [...] On social media, they can find people who share their identities and interests".[5] In other words, despite the inherent scepticism towards information quality on these platforms, users continue to engage, driven by a deep-seated desire for social connectivity and identity affirmation.[6] Even amidst widespread misinformation, the fundamental need for connection prevails, influencing user engagement intensity.[7] The research underscores the complexity of this interaction, noting that social media news consumption is tied to higher conspiracy beliefs, yet trust in social media news moderates this relationship, highlighting the critical role of misinformation identification.[8] Lastly, research on the demographic nuances of who trusts social media, found that women and younger users exhibit higher expectations of integrity, further emphasizing the social and empathetic dimensions of online engagement.[9]

Open-mindedness (N = 721)

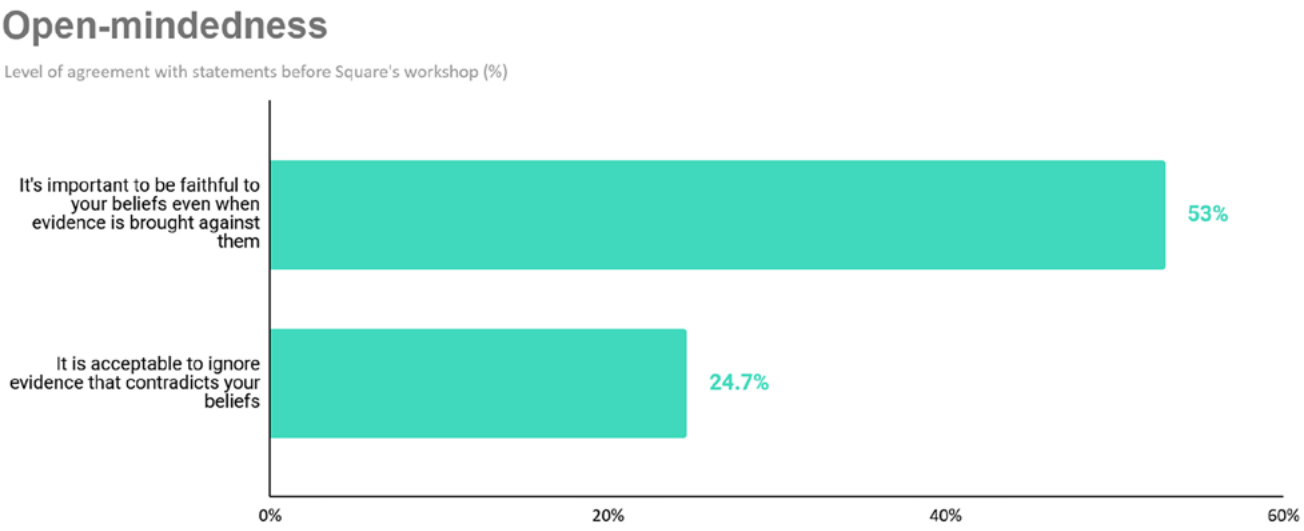
"Open mindedness refers to 'the tendency to evaluate new evidence in relation to a favoured belief, to devote sufficient time to a problem before giving up, and to consider the opinions of others carefully before forming one's own'.[10]
Less open-minded people are more likely to believe in conspiracy theories, are less good at discerning true information from false, and tend to be more susceptible to pseudo-profound statements.[11]
More open-minded people are more likely to revise their false beliefs; they are also more likely to reconsider their conspiracy beliefs when exposed to contradictory information".[12]

Sacha Altay

Open-mindedness, as quantified by the Actively Open-minded Thinking (AOT) scale[13], is argued to be essential in decision-making, belief evaluation, and evidence assessment. This psychological construct encourages the examination of different viewpoints, challenges personal beliefs, and adapts thinking based on new information, which is vital for sifting through complex information environments. Research has demonstrated that open-mindedness is associated with improved decision-making and reduced cognitive biases, fostering a more nuanced processing of new information and reducing biases, judgment errors, and dogmatism.[14]

Accordingly, Square's workshop incorporated pre- and post-measures of AOT to evaluate its effect on participants' open-mindedness.

Our cohort shows resistance to changing their beliefs even when faced with contradictory evidence. Only 27% seek out viewpoints that differ from their own. Participants were also asked about their agreement with two statements related to open-mindedness. The results suggest a tension between valuing evidence in principle and the reluctance to adjust personal beliefs when confronted with new information.

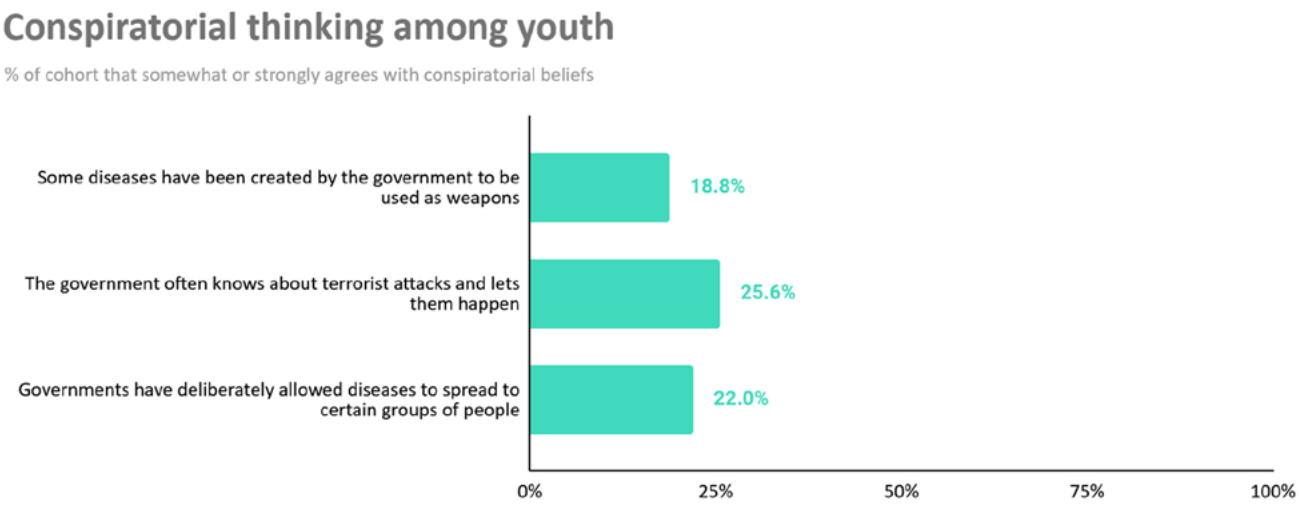


Conspiratorial thinking (N = 313)

Conspiratorial thinking matters significantly because it shapes individuals' perceptions of reality, influences their decision-making processes, and can profoundly affect societal trust, public health initiatives, and democratic institutions.[15]

The development and validation of the Adolescent Conspiracy Beliefs Questionnaire (ACBQ)[16], aimed at measuring adolescents' beliefs in conspiracy theories, highlight the importance of understanding the psychological antecedents and consequences of conspiracy thinking in young populations. This is particularly relevant for designing interventions that can effectively counteract the spread of conspiracy theories and misinformation among adolescents.[17] Thus, understanding these individual differences is key to developing informational interventions that can target these underlying factors, thereby inciting meaningful changes in thinking and building resilience against the allure of conspiracy theories.[18]

Our findings reveal that, on average, 22.2% of respondents endorse conspiratorial beliefs. These results highlight significant scepticism toward institutional motives, which aligns with broader trends of low institutional trust and limited civic confidence within the cohort. Conspiratorial thinking has far-reaching consequences, shaping individuals' perceptions of reality, influencing decision-making, and undermining trust in public institutions, health initiatives, and democratic processes.[19] Among young people, these beliefs can reinforce civic disengagement, scepticism toward credible sources, and increased susceptibility to misinformation. While providing valuable evidence within this demographic, the study's focus on a singular geographical and socio-economic context with specific age-ranges warrants caution when extrapolating these findings; comparative studies across diverse populations would prove essential for a comprehensive understanding of conspiratorial thinking at large.



Verification habits (N = 726)

This reluctance to engage with evidence that contradicts their beliefs is further emphasised by the gaps in critical behaviours. Over half of the cohort (61,5%) expressed confidence in identifying falsehoods yet lacked the habits needed to verify and challenge what they encounter online.

- Only 37% report checking the sources of the information they read.
- When asked to name verification tools, responses overwhelmingly relied on Google, AI tools (like ChatGPT), and social media comments, with little mention of structured fact-checking methods such as cross-referencing multiple sources or evaluating credibility.

These gaps suggest that while young people may recognise the questionable credibility of online content, they often lack the tools to critically assess and navigate it, leaving them vulnerable to manipulation and misinformation.[20]

[20] *Further qualitative analysis on middle-schoolers' behaviours and attitudes when facing online content can be found in Annex 1.*

Ability to distinguish facts from opinions
(N = 717)

The ability to effectively distinguish between statements of fact and expressions of opinion is a key component of critical thinking and media literacy. Participants were presented with a range of statements, spanning various topics from sports and personal preferences to social and scientific assertions, to assess their ability to make the distinction.
While certain statements were consistently categorized, others revealed notable variability in responses, revealing potential challenges in discerning subjective claims from objective truths. Notably, statements involving personal experiences, social commentary, or nuanced comparisons often led to increased uncertainty and a lower rate of accurate classification. These findings highlight the complexity of opinion identification and underscore the need for focused educational efforts to strengthen this essential media literacy skill.

Statement	Correct identification	% of uncertain
Messi is the best player in the world.	78.80%	5.0%
Emmanuel Macron does not care enough about climate change.	52.40%	11.0%
You're funny.	51.40%	7.5%
Fortnite is the best game in the world?	69.60%	9.1%
The Pacific Ocean is the largest ocean in the world	56%	21.0%
Teachers should give less homework	41.10%	12.0%
The earth is round.	77.6%	10.2%
Math is harder than French.	58.80%	11.5%
The school principal is strict.	71.60%	5.0%
It's too hot in France in summer.	32.50%	5.0%
Average	58.98%	9.7%

This table contrasts the evaluation of ten statements, distinguishing between factual and opinion-based responses before taking part in Square's workshop. Items evolved throughout the project, the number of answers ranges from 176 to 717.

Knowledge of the information landscape
(N = 640)

Understanding how different types of information function within the media ecosystem is essential for navigating digital content. To assess participants' conceptual knowledge, they were asked to define key concepts and to classify true and false statements related to digital governance, including free speech laws, platform regulations, and algorithmic filtering. Results suggest widespread uncertainty. Many respondents misjudged the limits of online expression and showed uncertainty about how social media algorithms filter content. Furthermore, 60% struggled to correctly define disinformation before the workshop, and only 23.5% understood confirmation bias. These findings highlight a lack of foundational knowledge about digital governance and suggest a need for more comprehensive digital literacy education that explicitly covers legal rights, platform governance, and algorithmic influence.

Statement	% True	% False	% I don't know
Freedom of expression is guaranteed by law in France.	71.6%	11.1%	17.3%
Insults about origins, sexual orientation and religion are prohibited by law and by the platforms.	65.9%	17.1%	17.1%
We're free to say whatever we like on social networks.	26.6%	63.6%	9.8%
Social media are designed to make us dependent on them.	30.0%	32.3%	37.8%
Algorithms filter the information displayed on our social networks.	35.2%	19.9%	45.0%

This table contrasts the evaluation of five statements before taking part in Square's workshop. Items evolved throughout the project; the number of answers ranges from N = 101 to 640.

Ability to assess information veracity

To complement our assessment of perceived online literacy skills, we evaluated middle schoolers' ability to assess veracity using a series of psychometrically validated misinformation items and real-world social media content. The stimuli ranged from text-based headlines and statements to more complex social media screenshots, incorporating engagement metrics, comments, and source information. This approach allowed us to assess not only their ability to evaluate isolated factual claims but also their susceptibility to misinformation in a more dynamic, digital environment that mirrors their everyday online interactions.

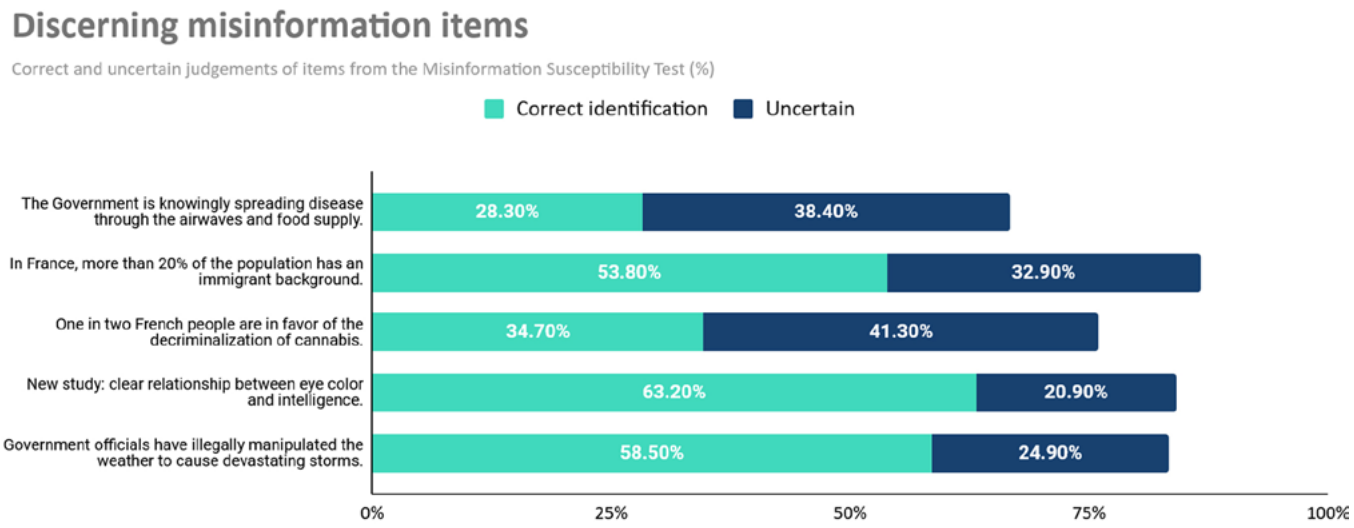
Example 1: Localised MIST item for veracity assessments.

Discerning veracity of psychometrically validated misinformation items (N = 559)

We presented them with a series of statements drawn from the Misinformation Susceptibility Test (MIST). This assessment applies a misinformation susceptibility test (MIST) scale, adapted to include locally relevant claims reflecting the political, media, and scientific narratives shaping young adults' information environments (see Example 1). A critical departure from the original scales is that we introduced an "I don't know" response option, which dominates across multiple items. However, some research suggests that the inclusion of «I don't know» options can negatively impact data quality, potentially leading to less valid responses.[21] Despite this, we believe it was essential to include this option here, as it allowed participants to express uncertainty rather than being forced into a false dichotomy, leading to a more nuanced understanding of their knowledge gaps. To our knowledge, this is the first application of such an option in MIST, aiming to avoid forcing participants into a dichotomous choice.

Our cohort shows random veracity assessment of manipulative content. On average, fewer than half of the responses correctly identified the truthfulness of statements, with uncertainty remaining high. This suggests a lack of confidence in assessing misinformation and points to the possibility of chance-level performance in evaluating these statements. This aligns with previous findings on foundational knowledge of information environments and verification strategies, highlighting significant gaps in critical evaluation skills.

However, the high level of uncertainty at baseline stands in contrast to the common assumption that young people overestimate their digital literacy. Rather than indicating sheer ignorance, this uncertainty could be interpreted as a form of desirable humility, a recognition of their own limitations in knowledge rather than unwarranted confidence in false beliefs.



Notably, in most cases, the combined proportion of correct identification and uncertainty outweighs outright incorrect answers. This suggests that while participants struggle with misinformation, they are not necessarily confidently misinformed, but rather hesitant and aware of their own uncertainty. While some statements may have been misleading by design, the high rate of misclassification suggests an ongoing difficulty in discerning truthfulness, highlighting the need for more targeted education on verification strategies and digital governance.

However, a key limitation of this assessment is that the statements were not presented in the standardised order of existing MIST scales (8, 16, or 21-item versions). This deviation from the original structure may have influenced how participants engaged with the task, potentially affecting their ability to compare statements or recognise patterns in misinformation. Additionally, instead of classifying statements as strictly real or fake, as in the original MIST measure, participants rated them on a scale from 100% false to 100% true. This shift moves away from a binary judgment of misinformation detection and instead captures a more nuanced assessment of perceived truthfulness, potentially reflecting uncertainty, partial belief, or skepticism rather than outright misclassification.

Evaluating online information in context
(N = 615)

The ability to discern between true and false is considered a core skill in studies on 'fake news' and misinformation. While in the past some studies only measured the ability to identify false or misleading statements, this method has been strongly criticised.[22] According to this method, it would be enough to adopt a strategy such as 'I mark everything false' or 'I only use the bottom of the scale' to obtain a perfect score! Today, most of the work aimed at reducing the spread of false information and improving critical thinking skills uses a discernment measure. The most common method for measuring discernment is to expose participants to true and false statements and ask them to assess the reliability of these statements. Then, it is only a matter of differentiating between the perceived reliability of the true and false statements: if the true statements are perceived as more reliable than the false statements, the discernment score is positive, otherwise it is negative.

To increase the external validity of these measures, i.e. to make them resemble the content to which people are exposed in everyday life, it is common to use news headlines in a format resembling that of social media.[23] The most used format is that of Facebook posts, with an image, a title, a subtitle and a source.[24] The selection of news headlines considered to be true is done via mainstream newspapers, such as Le Monde. The selection of false information is mainly done via fact-checking websites.[25] Researchers try to select recent headlines, especially for real news, as they are more likely to 'expire' and become false over time than the other way around.[26]
These tools were developed for adults and have mainly been tested in the United States. While some questionnaires have been adapted for audiences with a low level of education, such as rural populations in India[27], few or no discernment tools have been validated for adolescents. This is problematic, because the information practices of adolescents differ from those of adults: for example, adolescents are much more familiar with TikTok, Instagram or Snapchat formats than with Facebook.

In this report, we used screenshots of social media posts to measure discernment. This choice allows us to measure it as closely as possible to their everyday experience of social media. We used a Likert scale to measure the perceived reliability of the posts. In addition, the teenagers could leave a comment to explain their choice. This feature, which is uncommon in the literature, is very important for a detailed understanding of the reception and comprehension of this content. It allows for a better understanding of the reasons for their choices and the information used to make them. From a methodological standpoint, the comments can be very important for interpreting the results. (cf annex 1) For example, many quantitative studies that leave no room for comments have shown that false information generates more anger than real ones. The logical conclusion that many have drawn is that anger motivates the sharing of false information. However, a study that allowed participants to explain their choice by leaving a comment showed that a significant proportion of participants feel anger when reading false information because they are annoyed by the falsity of the news.[28] While anger may indeed facilitate the sharing of false information, an analysis of participants' comments suggests that it may also limit the sharing. Without these comments, it is difficult to know why participants feel these emotions and how they may or may not be related to the sharing of false information.

Sacha Altay

Prior to Square's workshop, the middle schoolers assessed the reliability of ten screenshots taken from social media (for examples, see Visual 2). While other measures allow us to maintain a controlled design and isolate specific factors (e.g., vocabulary used in headlines), this task requires students to integrate multiple cues, such as source credibility, engagement metrics, and visual elements. As this is more closely aligned with how they encounter news in everyday life, it serves as a highly ecologically valid context for assessing veracity. Together, these different measures offer a more nuanced understanding of how middle schoolers make sense of and judge the reliability of information they find online.

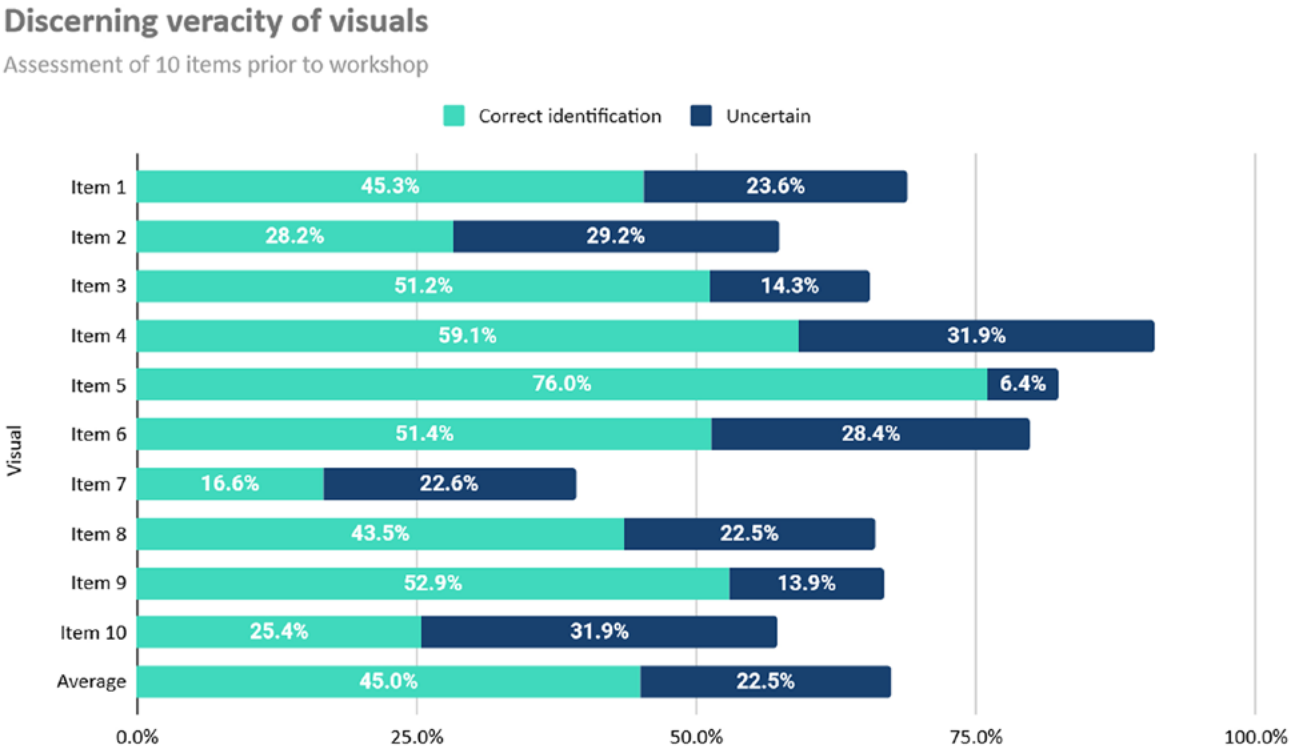
On average, 45% of responses correctly identified the visuals as true or false, while 22.5% expressed uncertainty by selecting «I don't know.» [29]

Example of social media screenshot that middle schoolers assessed

This finding underscores the challenges middle schoolers face in distinguishing reliable from misleading online content. With less than half of the responses accurately identifying the visuals, it suggests that students may lack the necessary skills or strategies to critically analyse digital media. The relatively high rate of uncertainty (average of 22.5%) reflects a lack of confidence but also suggests a degree of critical thinking as many students may recognise potential red flags and acknowledge their own limitations in verifying information which is a skill often underestimated. Indeed, incorrect assessments were less frequent than the combined total of correct and uncertain responses, suggesting that while students struggle with evaluation, they are not necessarily confidently misled but rather cautious in their judgments.

Further qualitative analysis on the screenshot and the middle-schoolers' attitudes and behaviours can be found on annex 1.

[29] *Items evolved throughout the project, the number of answers ranges from N = 222 to 615*



Section 2: Media literacy gaps in young adults

Sample insights

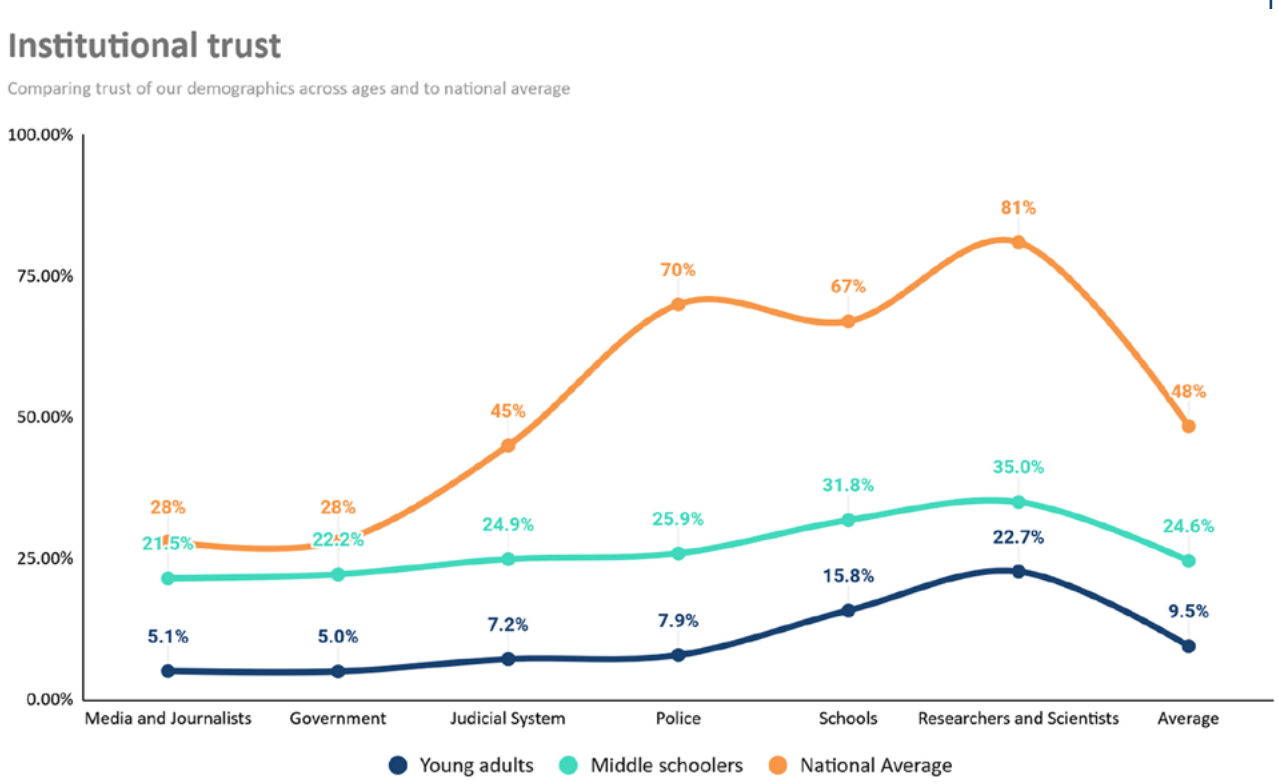
Our 630 young adults sample is balanced in terms of gender, with 52.3% male and 46.9% female, while 0.8% preferred not to answer. The age distribution ranges mostly from 18 to 25 years old, with the highest representation at 18 (22.2%) and 19 (18.2%). Most participants hold a Brevet des collèges (22.2%), followed by Bac Pro (13.3%) and Bac General or Technologique (11.9%), with higher education levels (Bac +2 and above) being less represented at 8.7%.

Institutional trust (N = 477)

Our results reveal a significant trust deficit. On average, young adults report significantly lower trust in institutions (9.5%) than the national average[30] in France. Our analyses suggest a strong negative correlation between government trust and conspiracy beliefs, meaning that lower trust in government is associated with higher endorsement of conspiratorial thinking.[31]

[30] *The national average spans across all ages. CEVIPOF- 2024 Barometer of Political Trust*

[31] *($r = -0.252$, $p < .001$)*

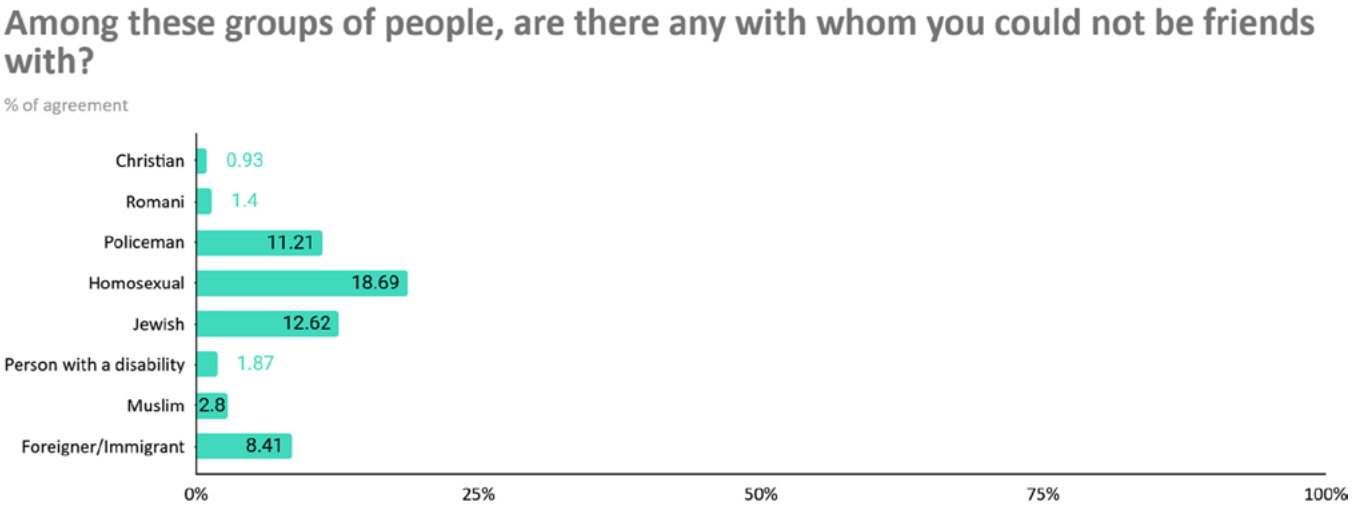


Tolerance for other social groups

(N = 214)

Nearly half of respondents do not feel they could be friends with all of the listed social groups.

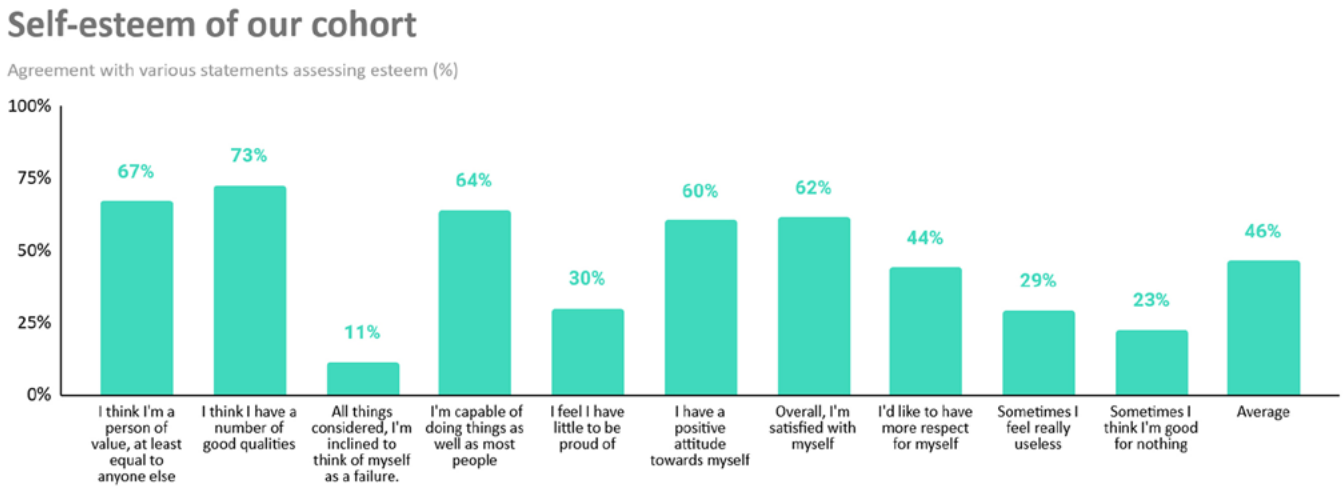
While only almost 54% of respondents indicated that they could be friends with all of the listed social groups, a significant minority expressed reluctance toward certain groups. Homosexual individuals and Jewish people were among those most frequently excluded, alongside police officers and immigrants. These responses highlight underlying social divisions and biases that persist within the cohort. Although rejection rates for some groups remain relatively low, the data suggests resistance toward individuals associated with law enforcement, LGBTQ+ identities, and religious or ethnic minorities. Future efforts should consider how to foster inclusivity, challenge prejudices, and promote social cohesion.



Self-esteem (N = 462)

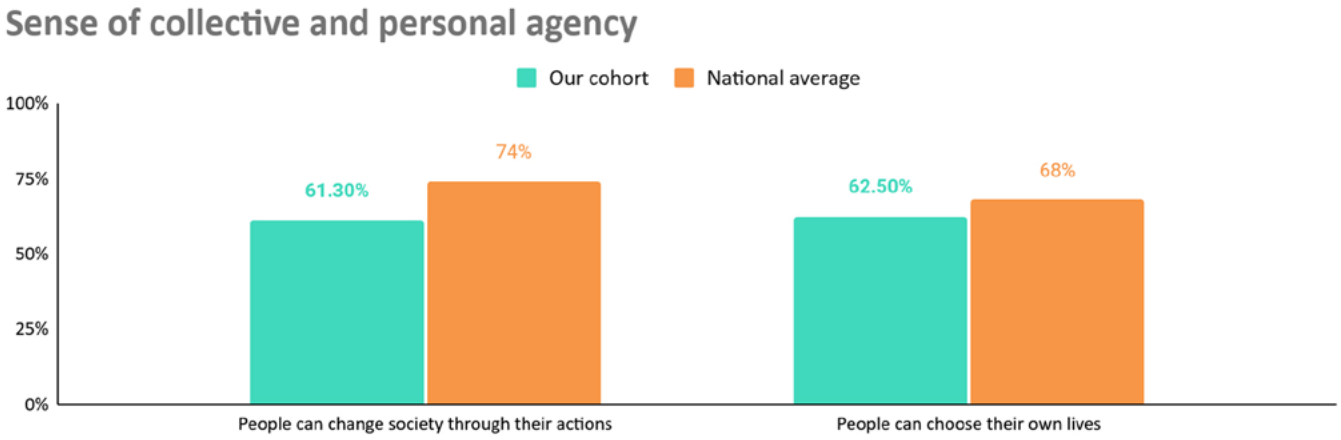
To understand the level of self-esteem, we used Rosenberg's self-esteem scale.[32] The results indicate a moderate level of self-esteem within the cohort, with stronger agreement on positive self-perceptions and notable variation in responses to more negative self-assessments. Most participants feel they have value and recognise their good qualities, suggesting a generally positive self-view. However, nearly 30% report feeling useless at times, and 22.6% believe they are «good for nothing», highlighting pockets of self-doubt and insecurity.

Interestingly, while a majority express self-worth and capability, a significant 44.1% wish they had more self-respect, suggesting that while they acknowledge their strengths, they may still struggle with self-acceptance or confidence. Overall, the cohort's average agreement with self-esteem statements is 46%, placing them in a moderate range. This suggests a balanced but somewhat fragile self-perception, where positive self-assessments coexist with moments of doubt and insecurity.



Sense of agency (N = 467)

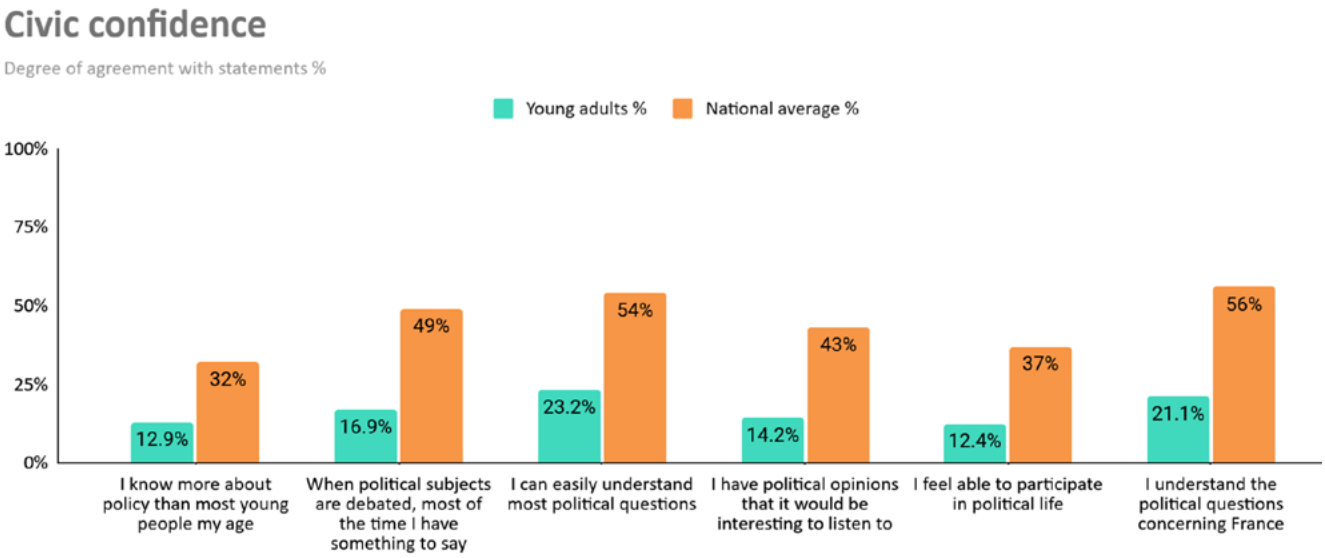
Our cohort demonstrates moderate confidence in both personal and collective agency, though their belief in the ability to change society through action is notably lower than the national trend. Their confidence in personal autonomy is closer to the national average[33], suggesting they feel more control over their own lives than over broader societal change. We must however consider that these national averages often focus on the general population (and are thus not limited to specific age ranges), making direct comparisons difficult. Still, this gap in collective agency may reflect socioeconomic barriers, institutional distrust, or limited civic engagement opportunities, which could impact their willingness to participate in public life. Since agency is crucial for media literacy and resisting manipulation, reinforcing civic participation and critical engagement skills could help strengthen their sense of empowerment and impact.



Civic confidence (N = 467)

Young adults display a worrying lack of civic confidence and consistently report feeling less knowledgeable and less engaged in political discussions compared to the national average.

A particularly stark gap appears in their ability to express political opinions and participate in debates, suggesting that many feel unprepared or excluded from political discourse. Furthermore, young adults struggle with understanding political questions at a much higher rate than the general population, which may contribute to their disengagement. This trend is consistent with data on middle schoolers, where civic trust is already low, and here too, we witness a further decline in confidence and engagement as young people grow older. These insights point to a broader issue of political alienation, underscoring the need for targeted efforts to enhance political education and engagement among young people.



Trust in social media and verification habits (N = 469)

Whereas 58.9% young adults report low or no confidence in the information they see on social media, it is still where they mainly get their information from, suggesting that they are not looking for the truth: a paradox consistent with the findings on middle-schoolers. Both cohorts evolve in an informational space they knowingly don't trust, suggesting that they might not be looking for trustworthy information. Furthermore, in young adults we found limited critical engagement with information. Only 26.3% actively check the sources of the information they read, and just 24.5% report engaging with viewpoints different from their own, suggesting a risk of echo chambers and passive consumption of information.

There are gaps in verification habits. Our cohort displays a heavy reliance on Google and general internet searches, while fact-checking services and traditional media remain underutilised. Few cite official sources, and some turn to non-traditional platforms like ChatGPT and TikTok, reflecting a lack of structured verification methods.[34]

[34] *N = 78*

Open-mindedness (N = 464)

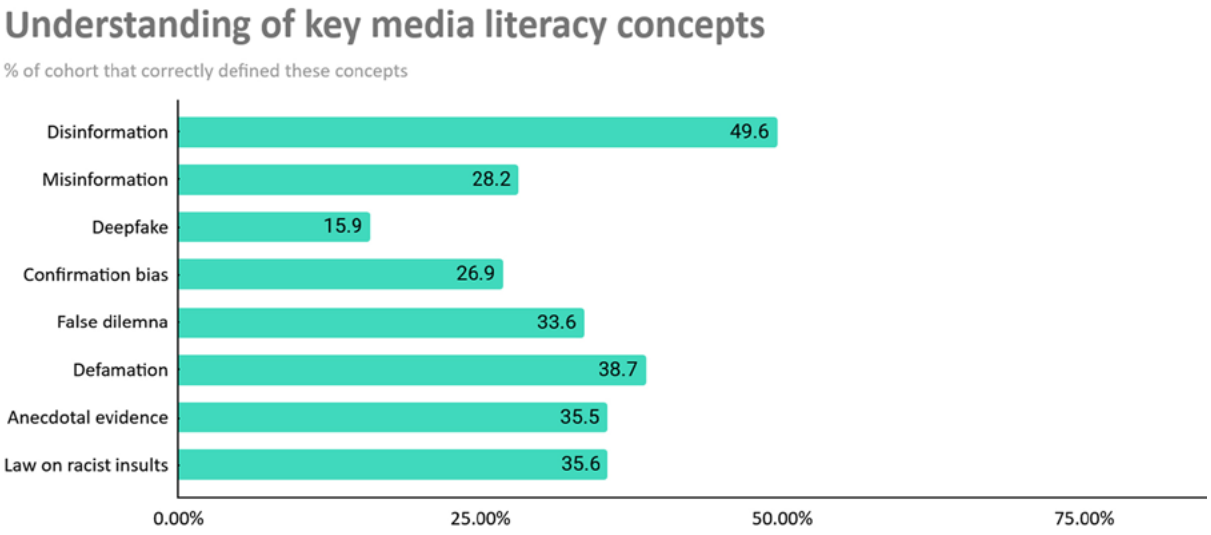
The data reveals a tension between openness to new information and resistance to change.

While many value reconsidering conclusions and exploring multiple perspectives, a significant portion prioritise conviction over adaptability, with 45.8% believing in holding onto beliefs despite contradicting evidence and 25.2% finding it acceptable to ignore counterevidence. Encouragingly, doubt is widely accepted, but the persistence of views favouring confidence over questioning emphasises the need to strengthen critical thinking skills.

Knowledge of the information landscape (N = 445)[35]

While almost 70% do not know what an algorithmic bubble is, less than 50% of the cohort correctly defined terms such as disinformation, misinformation, and deepfake pre-workshop.

[35] *Items evolved throughout the project; the number of answers ranges from N = 161 to 445 (average N = 253).*



Results show significant uncertainty, particularly regarding platform regulations and algorithmic influence. Only 30.9% correctly identified that platforms are legally required to participate in the fight against disinformation, while 58.3% expressed uncertainty. Similarly, while 51.6% correctly recognised that algorithms filter the information displayed on social media, 37.3% were unsure. Arguably, this uncertainty may reflect a more nuanced hesitation when dealing with complex digital concepts. Participants may be aware that these topics involve regulatory and technological complexities, making them less likely to provide confident but incorrect answers.

However, misjudging these digital governance principles can still affect their ability to critically engage with information online. Strengthening education on platform governance, algorithmic influence and digital rights could help provide the necessary foundations for navigating the modern information ecosystem more effectively.

Statements	% correct identification	% uncertain
The law stipulates that platforms must participate in the fight against disinformation.	30.90%	58.30%
Freedom of expression is guaranteed by law in France.	48.10%	29.30%
Insults about origins, sexual orientation and religion are prohibited by law and by the platforms.	52.80%	30.30%
We're free to say whatever we like on social media.	52.60%	21.30%
Social media are designed to make us dependent on them.	34.90%	37.70%
Algorithms filter the information displayed on our social media.	51.60%	37.30%
The argument from authority aims to convince by using the origin of a statement rather than its content.	23.80%	70.80%
Is the information most visible on social media the most reliable?	84%	-

Ability to distinguish facts from opinions (N = 459)[36]

Participants were asked to distinguish facts from opinion within different statements, covering topics from basic science to historical and political facts. While straightforward, widely known facts, such as the number of letters in the alphabet, were more likely to be correctly identified, responses indicate greater uncertainty in areas that require deeper contextual knowledge.

For example, only 48.5% correctly stated that water boils at 100 degrees, while the lowest correct identification rate was for the statement that France is a secular republic and, within this framework, the law does not recognise blasphemy and offence, with just 25.1% answering correctly and 40.7% uncertain. While participants show confidence in widely accepted factual knowledge, they are noticeably more cautious when assessing statements related to geopolitics, governance or legal principles. This uncertainty may reflect intellectual caution and, at times, identity-based reasoning, rather than outright misinformation, suggesting an awareness of their own knowledge limitations.

[36] *Items evolved throughout the project; the number of answers ranges from N = 221 to 459 (average N = 314)*

Statement	% correct identification	% uncertain
You don't have the right to blasphemy when you don't believe in anything!	47.30%	37.40%
Messi is the best player in the world.	61.40%	15.80%
Emmanuel Macron does not care enough about climate change.	30.70%	29%
In France, it's hotter in summer than in winter.	71.80%	15%
Water boils at 100 degrees.	48.50%	39.50%
You're funny.	44.60%	16%
It's too cold in winter in France.	27.70%	13.60%
There's no real problem of police violence in France.	64.70%	22%
"They cheated; I won." (Trump after Joe Biden's election victory.)	42.40%	40.10%
During the Yellow Vest protests in France between November 2018 and March 2019 the government announced 2,200 protesters injured.	45.90%	43.60%
Russia launched an attack on Ukraine on 24 Feb 2022.	63.10%	27.50%
France is a secular republic and within this framework the law ignores effectively the offence of blasphemy.	25.10%	40.70%
Russia attacked Ukraine to prevent its accession to NATO.	33.10%	41.60%
There are 26 letters in the alphabet.	78.30%	15.20%

Conspiratorial thinking (N = 459)

Conspiratorial thinking is a psychological disposition to believe that events or situations are the result of secret plots orchestrated, in general, by powerful and malicious groups.[37] This dimension is particularly useful for statistical purposes because it has a normal distribution: most people are moderately conspiracy-minded according to this scale, and some are more so than others. This contrasts with adherence to specific conspiracy theories, particularly the less popular ones such as the flat earth, which can have a bimodal and asymmetrical distribution: a majority does not believe in them at all, and a small number believes in them strongly. Conspiratorial thinking correlates rather strongly with adherence to conspiracy theories.

Sacha Altay

Mistrust runs deep, with 69.5% agreeing that the government is deliberately hiding important information from the public, reinforcing a broader pattern of institutional distrust.

On average, almost 40% of our cohort agree with conspiratorial beliefs, with many believing that authorities withhold information, engage in surveillance, and allow harmful events to unfold deliberately. Public health conspiracies, such as governments weaponizing diseases, also see notable support. While extreme theories, such as 5G microchips in vaccines, remain less common, scepticism towards major political and historical events, including election fraud and state-orchestrated attacks, suggests a broader pattern of institutional distrust. This trend underscores the importance of early interventions in media literacy and critical thinking, before these distrust-based narratives become deeply ingrained. Equally, we need qualitative interventions that can meet older groups where they are, acknowledging their existing distrust while equipping them with critical thinking skills and media literacy tools to navigate misinformation effectively.

Statement	% Agreement
The government is deliberately hiding important information from the public.	69.50%
The government is secretly monitoring people.	55.70%
Some political groups have secret plans that are not good for society.	63.70%
Some diseases have been created by the government to be used as weapons.	40.70%
The government often knows about terrorist attacks and lets them happen.	29.60%
Governments have deliberately allowed diseases to spread to certain groups of people.	40.50%
Secret groups control people's minds without their knowledge.	30%
Secret societies control politicians and other leaders.	40.50%
Secret societies influence many political decisions.	49.90%
The Covid-19 vaccine contains 5G microchips to track us.	10.60%
The American government lied to justify the war in Iraq.	42.10%
It is certain that the Charlie Hebdo attacks were planned and carried out by the government.	50.30%
Experts talk about climate change to limit economic development.	24.60%
The result of the 2020 U.S. presidential election was rigged from the start.	23.40%
A vast network of corruption and money laundering exists within the government.	17.30%

Note: This table includes all conspiratorial statements tested over recent years. The first 9 statements are from the ACBQ and have up to 450 responses. The last six statements were discontinued, so sample sizes here are up to 233.

Actual abilities to discern veracity across different formats

Discerning veracity of psychometrically validated misinformation items (N = 187)

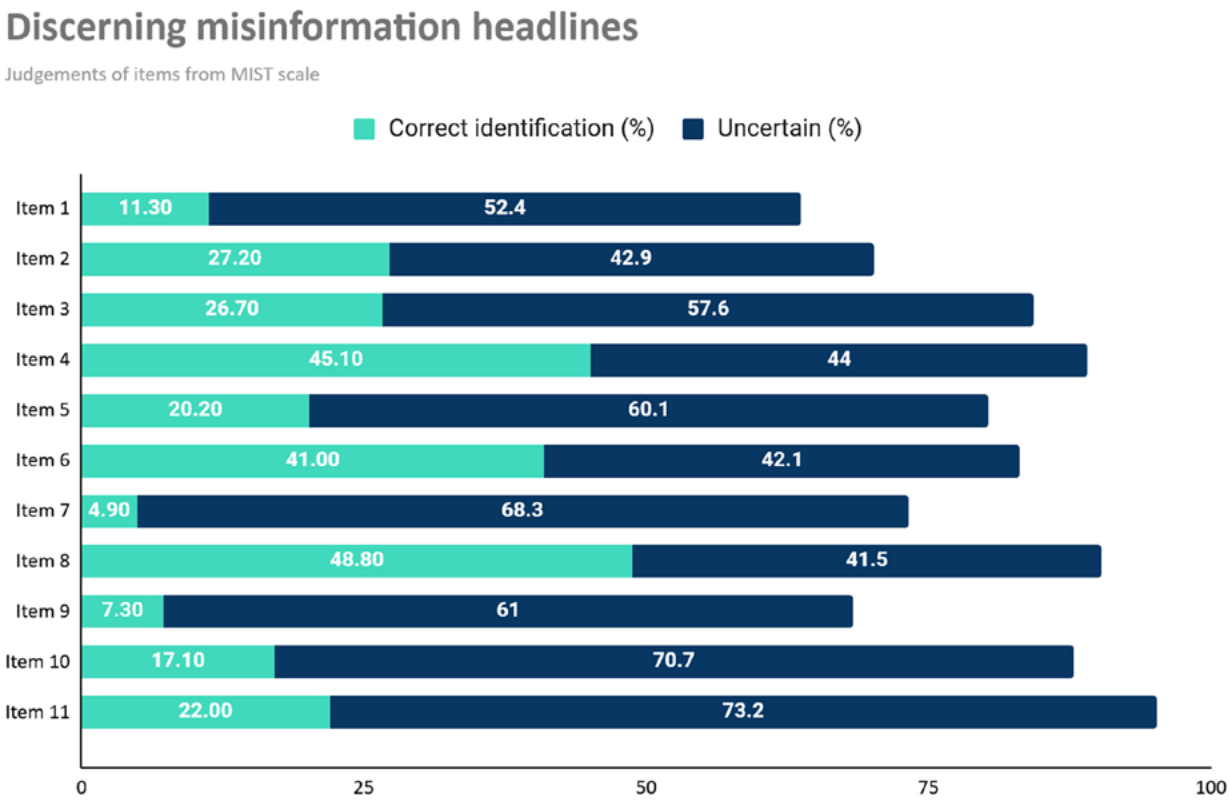
The results suggest that misinformation susceptibility in this cohort is less about outright belief and more about uncertainty.

Rather than confidently accepting falsehoods, many participants grapple with assessing veracity, demonstrating uncertainty in evaluating information rather than a firm belief in misinformation. While this highlights significant gaps in knowledge, it also suggests a form of intellectual humility, an awareness of their own limitations rather than unwarranted confidence in incorrect information. This openness to doubt creates opportunities for further learning, as it is easier to equip individuals who recognise their uncertainty than to challenge deeply entrenched false beliefs.

Specifically, political and institutional claims, such as "The government is manipulating the public's perception of genetic engineering" (Item 1) and "The corporate media is controlled by the military-industrial complex: the major oil companies own the media and control their agenda" (Item 10), see low correct identification rates and high uncertainty. Similarly, health-related misinformation, such as "The government is knowingly spreading disease through the airwaves and food supply" (Item 2), sees low accuracy rates, pointing to potential vulnerabilities in public health misinformation discourse.

Rather than clear-cut misinformation belief, this cohort's susceptibility is shaped by uncertainty and hesitation.

The lack of confident rejection of false claims suggests that many young adults may be persuadable when faced with misleading narratives, making them more susceptible to misinformation influence over time. The accompanying graph illustrates how misinformation susceptibility manifests not just in belief, but in doubt.



Note: This set of MIST items was not presented in the traditional 8-, 16-, or 21-item scale format and is therefore primarily exploratory. Some items were replaced to better fit the context, and as a result, the sample size varies across statements, ranging from 42 to 187 participants (average N = 114).

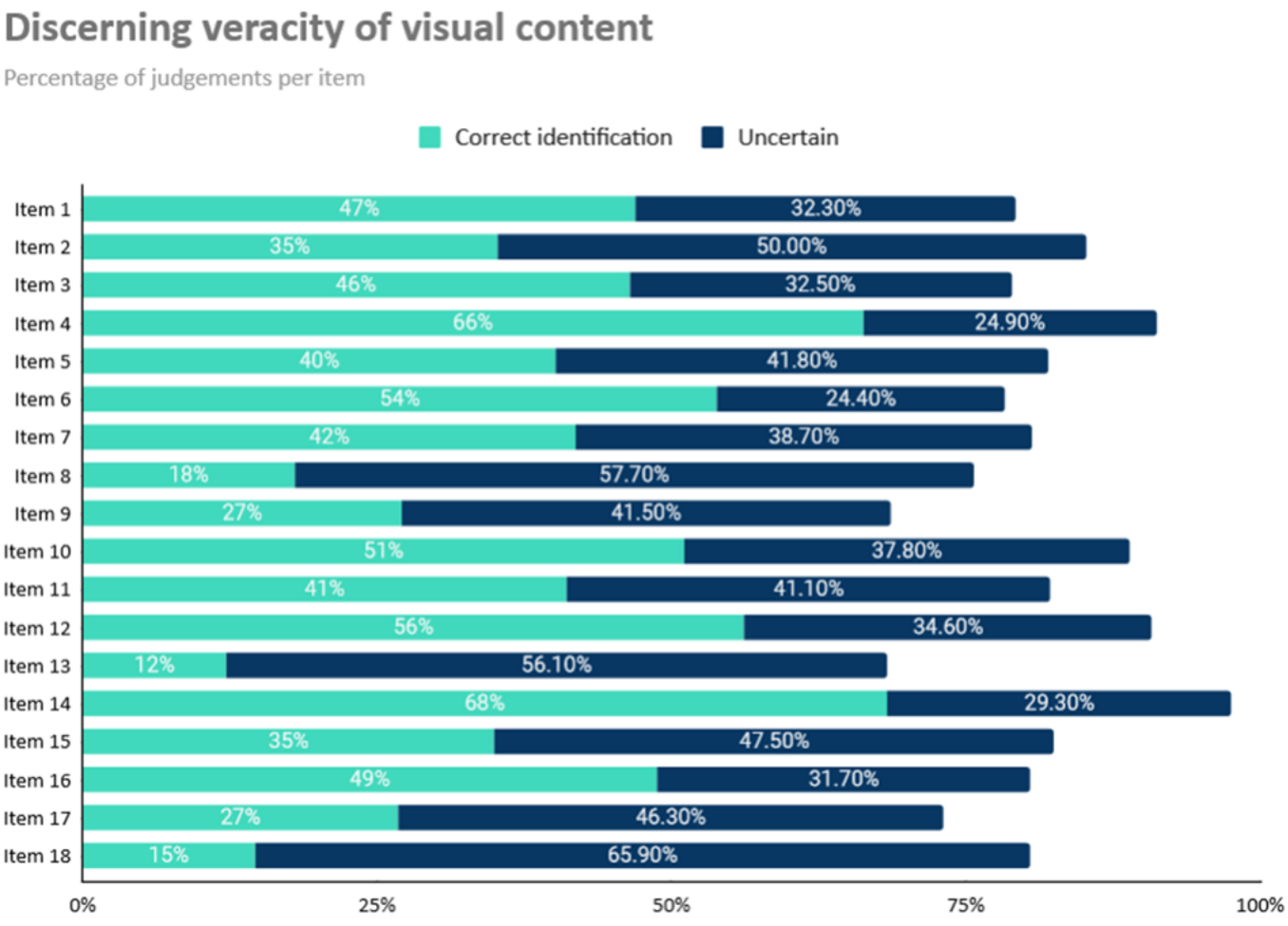
Evaluating online information in context (N = 200)

Across the workshops, young adults assessed 18 screenshot-based items from social media to gauge their ability to judge visual content reliability. The data reveals significant uncertainty, with uncertain responses often matching or exceeding correct identifications.

On average, only 41% of responses correctly classified the visuals, while uncertainty surpassed 50% on several items.

This suggests that many young adults struggle to evaluate credibility, as assessing source reliability, engagement metrics, and visual cues requires critical analysis skills they may lack.

The high uncertainty rate signals a lack of confidence in verification, where participants may recognise potential red flags but struggle to determine accuracy or seek reliable confirmation methods.



Note: Some items were replaced to better fit the context, and as a result, the sample size varies across statements, ranging from 41 to 200 participants (average N = 149).

Conclusion:
Bridging the media literacy gap

Our findings reveal deep-rooted vulnerabilities in the ability of both middle schoolers and young adults to navigate today's complex information landscape. Across key measures, including institutional trust, civic confidence, media literacy concepts, and misinformation susceptibility, our participants demonstrate low accuracy, high uncertainty, and weak verification strategies, leaving them exposed to manipulation.

The combination of institutional distrust, high levels of uncertainty and conspiratorial thinking suggests that many feel disconnected from civic life, question official narratives, and perceive themselves as excluded from meaningful discourse.

Our cohorts also demonstrate low levels of both foundational knowledge and understanding of the digital information environment, which are essential for assessing the credibility of digital content and identifying misinformation. In both groups, misinformation susceptibility is characterised more by uncertainty than outright belief, with many respondents selecting «I don't know» rather than confidently endorsing or rejecting false claims. This hesitation, especially in politically charged and digital misinformation contexts, suggests a lack of confidence in assessing veracity rather than fixed misperceptions. Moreover, verification habits remain weak. Young adults rely heavily on Google, social media, and AI-driven tools like ChatGPT, with little engagement with structured fact-checking methods or critical source evaluation. However, this uncertainty also presents an educational opportunity. Those aware of their own limitations, as seen in their expressed uncertainty, may be more open to learning new verification strategies. Strengthening digital literacy, particularly in source evaluation, fact-checking, and algorithmic awareness, can help equip young adults to navigate online information more critically and effectively.

Simply raising awareness is not enough. Addressing misinformation susceptibility and media illiteracy requires a structured, hands-on approach that prioritises skill development over passive learning. Our findings underscore the need for evidence-based interventions that foster critical thinking, teach verification strategies, and integrate prebunking techniques to counteract manipulation online.

Part II: Developing critical thinking

Square's workshops

The workshops aim to develop critical thinking skills, to support the ability to analyse information and to foster civic engagement of the participants. With research-backed experience, the training resources are evolving to offer activities tailored to our audiences and to what they see online. Square is constantly updating a bank of examples adapted to their interests as well as 'active' exercises to stimulate participants' engagement and foster interaction.

Square's approach combines 'prebunking' methods (where participants are put in a situation where they are confronted with manipulation techniques to learn to identify and resist them) with metacognition. Participants are supported with tools and activities that allow them to take action. Key topics include online manipulation techniques, freedom of speech, verification tools, distinguishing facts from opinions, understanding the context surrounding online disinformation and discussing biases.

Depending on context, Square offered workshops ranging from 8 to 20 hours over the course of 1,5 to 4 days for young adults. The above-described core content remained the same, but more time allowed for more in-depth exploration of themes for some groups and opportunities to create social media content. For middle school students, workshops took place during class hours (55 minutes), during 1 to 4 hours.

Evaluation design

To evaluate Square's workshop's impact on the perceptions and knowledge of young people, we conducted pre- and post-workshop questionnaires. The end-of-workshop questionnaire mirrored the one given at the start, covering the same dimensions of interest. This allows for a before-and-after comparison for each dimension. This method is based on the assumption that young people's perceptions and knowledge would have remained the same if there had been no workshops. As the time between the two questionnaires is short, this reduces the probability that events will significantly alter young people's perceptions and knowledge outside the workshops.

Follow-up data collected 6 to 12 weeks after the workshop is presented for completeness, but attrition is very high, which reduces the scope of the analyses.

Some evaluation items changed during the project - one of the aims of which is precisely to test evaluation items, complicating comparisons across testing waves.

Section 1: Middle schoolers

The following section presents insights into students' knowledge retention, reasoning abilities, and behavioural shifts, but the high attrition linked to lower participation in post-test (N = 180 VS N = 405 at pre-test[38]) and follow-up (up to 75) considerably limits the significance of the results. The main aim of assessing skills at this stage was to identify relevant items and establish the above needs analysis. A rigorous evaluation of the impact of workshops designed based on the analysis of these needs is planned for the next phase of the project.

[38] *The pre-test sample is smaller than the one analysed above for the diagnosis (N = 751) because only those middle school pupils who took part in a workshop after completing the pre-test are included here, to enable a comparison to be made between pre and post-test responses. Yet half of the sample analysed in the first part was only consulted for the diagnosis and did not receive any intervention.*

Definitional knowledge of key media literacy concepts

We assessed students' understanding of disinformation and confirmation bias at three time points: before, immediately after, and up to three months after the workshop. The findings reveal both progress and ongoing challenges in conceptual learning.

For *disinformation*, correct identification initially stood at 39.6% (N = 398), dropped at post-test (N = 127), but then rebounded significantly at follow-up (N = 46), suggesting a delayed learning effect. Notably, uncertainty, which was high at pre-test, dropped significantly at post-test and remained relatively low at follow-up. This indicates that while some conceptual challenges remained, participants were more decisive in their responses over time, reducing the tendency to respond with "I don't know".

For confirmation bias, correct responses increased significantly from after taking part in the workshop, demonstrating clear improvement in understanding (N = 401 at pre-test and N = 114 at post-test). At follow-up (N = 47), correct responses declined, and uncertainty rose sharply, suggesting that while initial learning gains were observed, retention remained inconsistent over time.

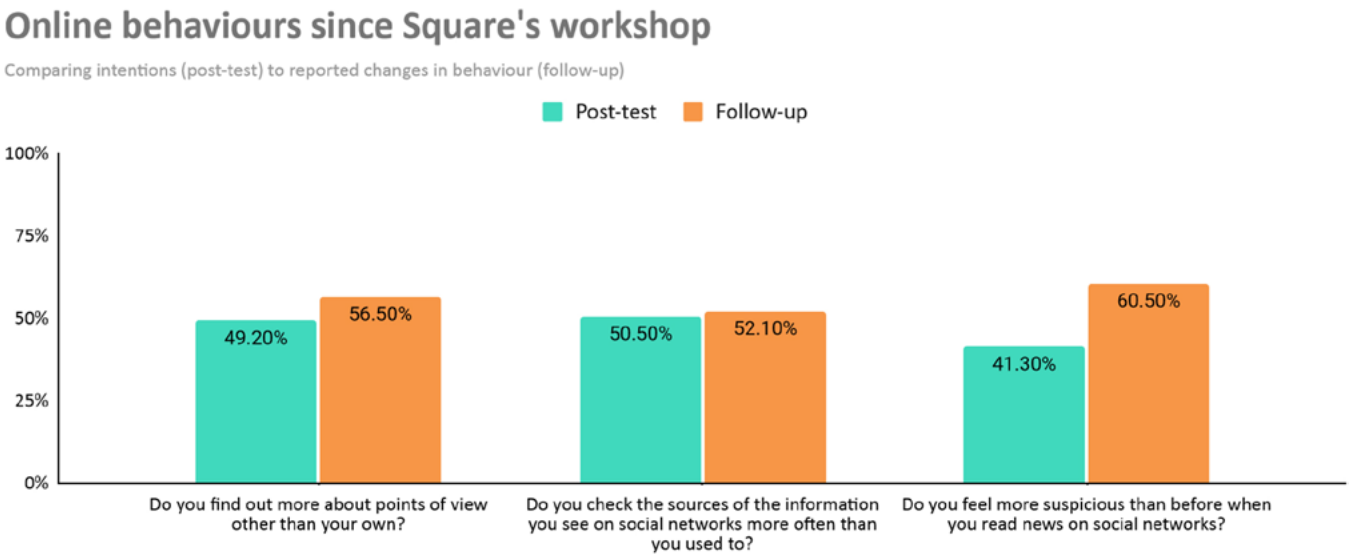
These findings highlight that participants demonstrated learning gains post-workshop, but retention varied. The stronger post-test improvements in confirmation bias suggest that certain concepts were more immediately grasped, whereas the delayed rebound in disinformation understanding raises important questions about how and when media literacy concepts solidify. Additionally, high sample attrition at post-test and follow-up limits the ability to fully assess long-term knowledge retention across the entire cohort. These results underscore the need for sustained reinforcement and more longitudinal follow-ups to determine how conceptual learning evolves beyond a single intervention.

Discerning between fact and opinion

Our analysis of students' ability to distinguish between factual statements and opinions reveals modest variations across different items and time points, with no significant differences between pre- (N = 176 to 400 depending on items) and post-test (N = 128 to 139) performance. Objective statements, such as «The earth is round», were more consistently identified as factual, while evaluative or subjective statements (e.g., «Messi is the best player in the world» and «The school principal is strict») showed lower accuracy and higher uncertainty. Indeed, subjective statements seem to continue being a challenge for participants even after the intervention.

Reported behavioural changes

At post-test (N = 138) and three months post-workshop (N = 48), participants report notable shifts in their online behaviours, with over half indicating that they will or that they now check sources more frequently and seek out diverse perspectives more than they did before. A majority also express increased suspicion toward news on social media, suggesting a heightened critical awareness of information reliability. These findings highlight positive movement toward more critical engagement with digital content, particularly in verification practices and exposure to different viewpoints. However, while these self-reported changes indicate progress, further research is needed to assess whether these behaviours are consistently applied and lead to improved accuracy in information discernment over time.

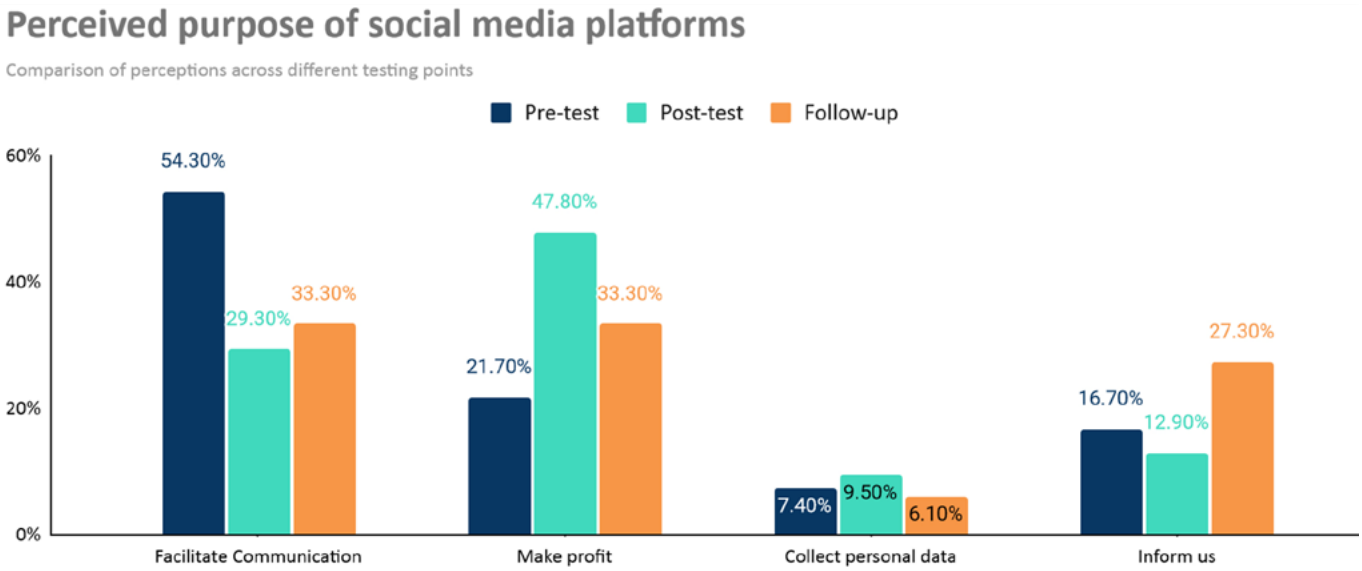


Section 2: Young adults out of school

Knowledge of key media literacy concepts, legal rights and digital regulations

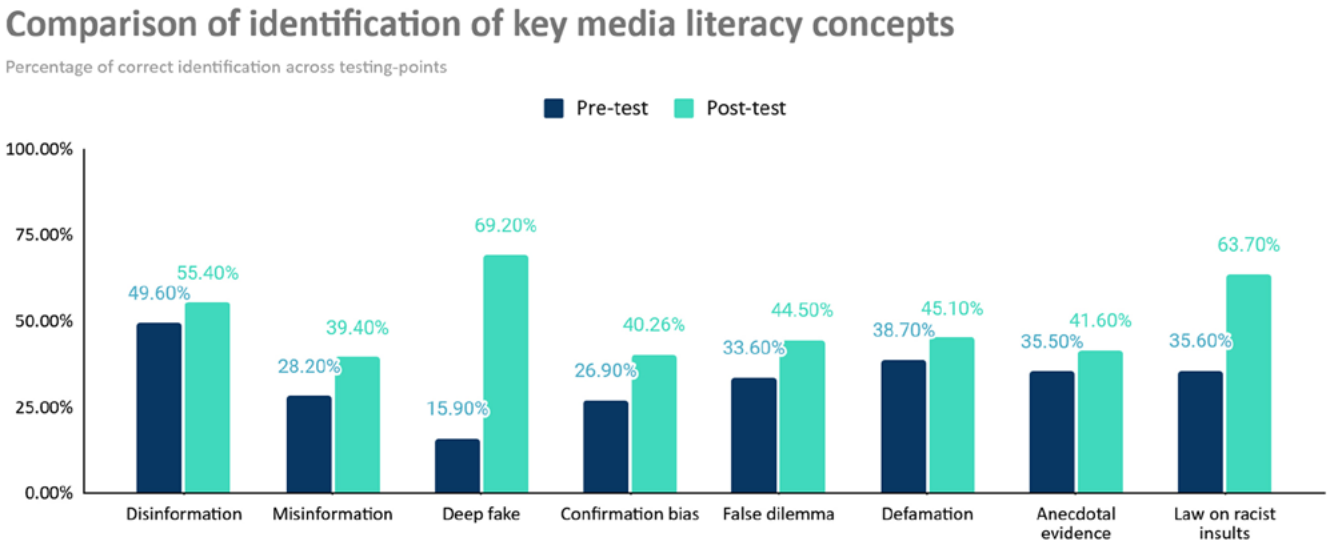
Our results suggest that after participating in Square's workshop, participants demonstrated significant improvements in their understanding of key aspects of legal rights and regulations and media literacy concepts.

Firstly, the results show a clear shift in how participants perceive social media's purpose. There was a notable increase in recognising its profit-driven nature (N = 233). Over time, perceptions partially reverted, though communication did not regain its initial dominance, and the view of social media as an informative tool increased (N = 66). These findings suggest that the workshop successfully broadened participants' understanding of social media's functions, but without reinforcement, some initial shifts may diminish over time. This underscores the importance of longitudinal efforts to ensure that awareness translates into sustained critical engagement rather than passive acceptance of information flows.



Secondly, our results suggest that after participating in Square's media literacy workshop, participants demonstrated significant improvements in their understanding of all[39] key media literacy concepts (N = 164 to 398 depending on items). Notably, these gains persisted over time, with no significant decline observed at follow-up testing, but a considerable imbalance in sample sizes, due to attrition, may affect the robustness of the follow-up data (N = 67).

[39] *The improvement in disinformation definition was not statistically significant (p = 0.126).*



Our cohort also showed significant improvements in key aspects of legal rights and regulations. Several items showed increases in correct responses from pre-test to post-test (N = 234 to 407 depending on items), paired with fewer "I don't know" answers, indicating that participants felt more confident in evaluating claims around law, free speech, and social media practices. The follow-up data presents a more mixed picture (N = 69[40]). Some statements, such as "*The argument from authority aims to convince by using the origin of a statement rather than its content*", maintained much of their post-test improvement, suggesting sustained learning. Others, particularly the one related to algorithmic filtering, saw a re-emergence of uncertainty at follow-up, in some cases exceeding pre-test levels. This could indicate that, despite initial gains, participants are finding these topics more complex than expected. The increase in uncertainty likely reflects the inherent difficulty of these issues, with participants feeling unsure or willing to admit to the complexity and lack of clarity around legal rights and platform regulations.

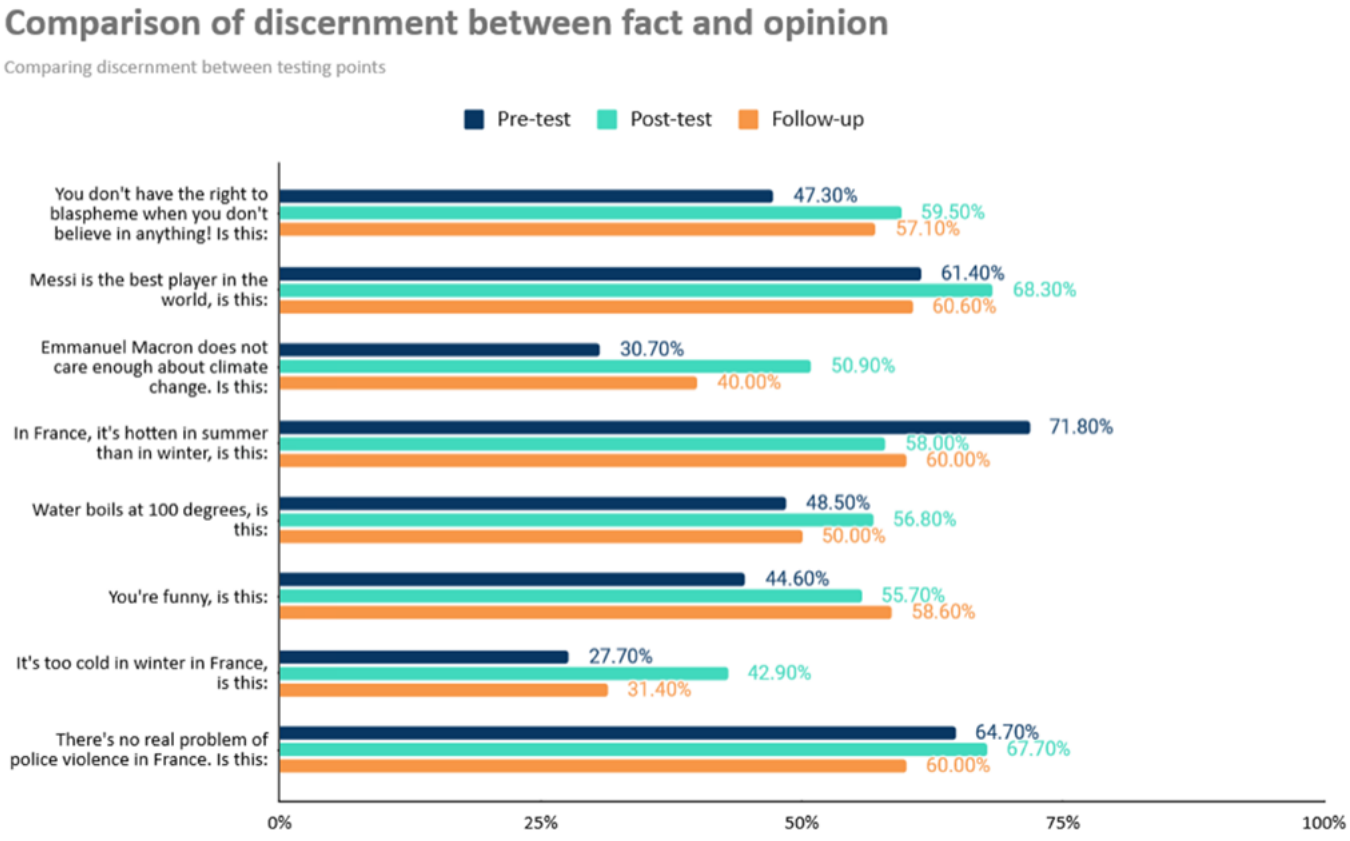
[40] *The follow-up sample was substantially smaller, reducing statistical power and potentially limiting the generalisability of long-term effects.*

Statement	Pre-test	Post-test	Follow-up	Pre-test	Post-test	Follow-up
		Correct identification (%)			Uncertain (%)	
The law stipulates that platforms must participate in the fight against disinformation.	30.90%	41.60%	39.10%	58.30%	41.60%	39.10%
Freedom of expression is guaranteed by law in France.	48.10%	56.20%	44.90%	29.30%	27.20%	34.80%
Insults about origins, sexual orientation and religion are prohibited by law and by the platforms.	52.80%	59.30%	53.60%	30.30%	20.30%	27.50%
We're free to say whatever we like on social media.	52.60%	52.10%	49.30%	21.30%	22%	37.70%
Social media are designed to make us dependent on them.	34.90%	45.50%	40.60%	37.70%	30.90%	34.80%
Algorithms filter the information displayed on our social media.	51.60%	63%	36.20%	37.30%	16%	42%
The argument from authority aims to convince by using the origin of a statement rather than its content.	23.80%	35.80%	35.30%	70.80%	47.40%	52.90%
Average	**42.10%**	**50.50%**	**42.71%**	**40.71%**	**29.34%**	**38.40%**

Effective media literacy initiatives must go beyond simply assessing knowledge, focusing on equipping individuals with critical thinking skills that can be applied across various contexts. In addition to exercises that allow the memorisation of answers, the workshops are designed to teach participants how to evaluate misinformation, navigate social media, and apply these skills to real-world situations where the line between truth and falsehood is often more blurred.

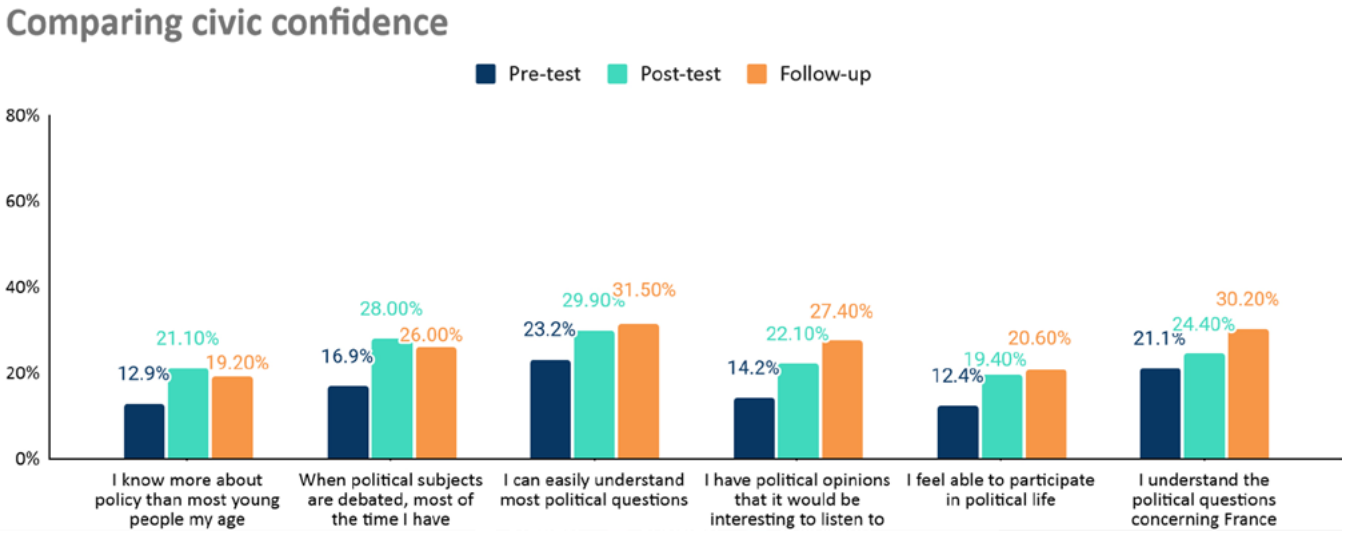
Differentiating between fact and opinion

After participating in Square's workshop, participants showed notable progress, with significant improvements across most statements (N = 175 to 392 depending on items). Encouragingly, many of these improvements persisted over time, suggesting that the intervention had a lasting impact for at least 3 months (N = 70).



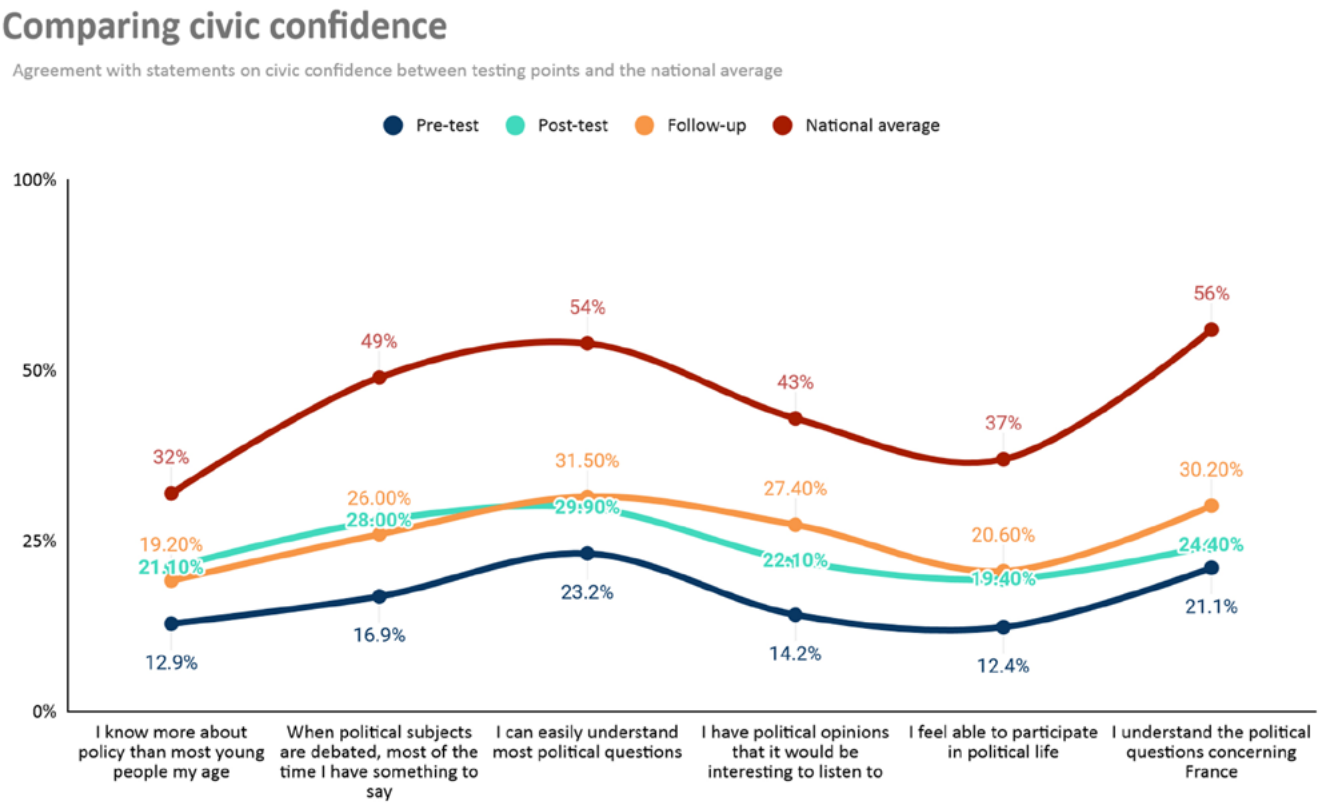
Civic confidence

Square's workshop fostered a sustained increase in political confidence (N = 355), with gains not only maintained but, in some cases, further strengthened up to three months later (N = 73 at follow-up), empowering participants to engage more actively in political discussions and express their views with confidence. Participants reported feeling more knowledgeable about policy, more engaged in political discussions, and more capable of understanding and contributing to civic life after participating in Square's workshop. Notably, the strongest gains occurred in political self-expression, with more participants agreeing that they have political opinions worth listening to and something to say in debates. This suggests that the intervention helped foster a stronger sense of political voice and engagement and maintain it for at least three months after taking part in the workshop.



However, despite these improvements, a significant gap remains between the cohort and national averages. Even at follow-up, confidence levels fall well below national benchmarks, particularly in understanding political questions (31.5% vs. 54%) and feeling able to participate in political life (20.6% vs. 37%).

These findings highlight that while progress has been made, deeper structural barriers to civic engagement persist, requiring long-term reinforcement and opportunities for practical political participation to close the gap.



Conspiratorial thinking

Average conspiratorial thinking scores showed a slight numerical decline, but our analyses found this change statistically insignificant. (N = 382)

In 2023, data showed a significant shift in attitude towards conspiratorial beliefs. Square's workshops lasted 20 hours over 4 days (against 2 days in 2024) and dedicated a specific session to conspiracy theories. The initial conspiracy level of this 2023 audience (recruited through local NGO channels) was much higher than the combined 2023/2024 ones (recruited by Mission Locales).[41]

Out of the 228 participants in 2023, we focused on 90 individuals who completed all the key measures (i.e., conspiratorial thinking, Actively Open-minded Thinking, fact vs opinion discernment, theoretical disinformation knowledge) and at least 85% of the remaining scales in our questionnaires (N = 90)[42]. With an initial level of 66.6% of participants registering some level of agreement with conspiratorial statements, this figure declined by 16.6 percentage points post-workshop. This substantial reduction was both statistically significant (p = 0.002) and had a notable effect size (Cohen's d = -0.6), showcasing the effectiveness of the workshop in altering beliefs in conspiracy theories. This is aligned with a systematic review that assessed the effectiveness of various interventions in countering conspiracy beliefs and found that interventions fostering an analytical mindset or teaching critical thinking skills were notably successful in altering such beliefs.[43]

[41] *66.6% of the participants registered some level of agreement with ACBQ conspiratorial statements in 2023 VS 40% in 2023/2024.*

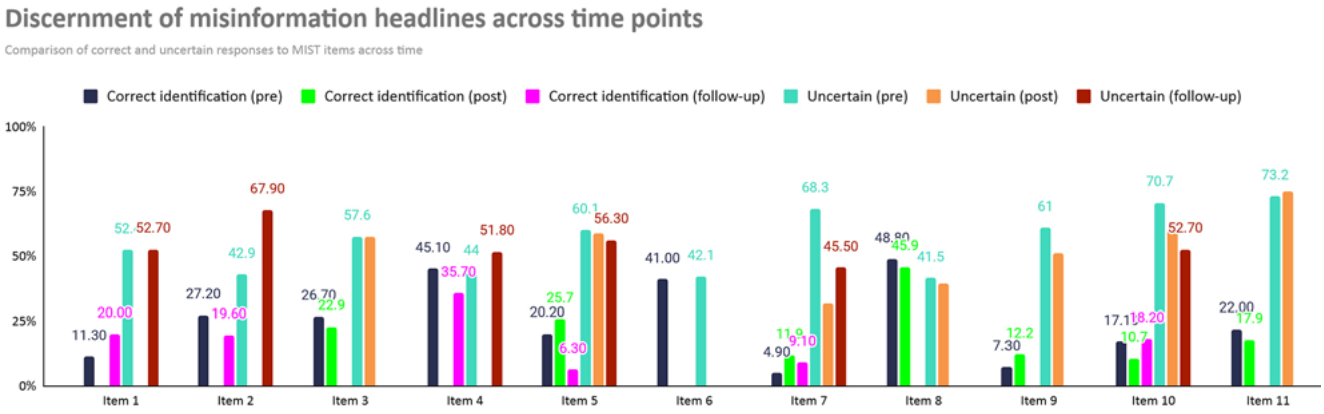
[42] *In a parallel analysis that did not apply a completion filter, Quentin Daviot and Simon Briole found that, on average, Square's workshops reduced conspiracy by 38% of a standard deviation, at the 1% significance level (N = 122). And that for young people who had a high level of conspiracy before starting the workshop, i.e. with a level of conspiracy above the median (N = 68), the workshop reduced conspiracy by an average of 1.08 standard deviations, at the 1% significance level. See annex 2.*

Actual abilities to discern veracity across different formats

The items were part of an exploratory endeavour, with different items introduced or removed throughout the workshops. We are still determining which items best assess impact on veracity discernment. Our findings highlight the unique challenge of measuring media literacy skills in dynamic, visually complex environments like social media.

Discerning veracity in misinformation headlines from the MIST

These data reveal a mixed and often incomplete pattern in participants' ability to discern the veracity of headlines over time (N = 29 to 174 at post-test). Some items show initial gains in correct identification from pre-test to post-test, while others remain stable or even decline. At follow-up, uncertainty resurges for several items, suggesting that without ongoing reinforcement, participants may lose confidence in evaluating misleading content (N = 17 to 57).



Assessing veracity in social media screenshots

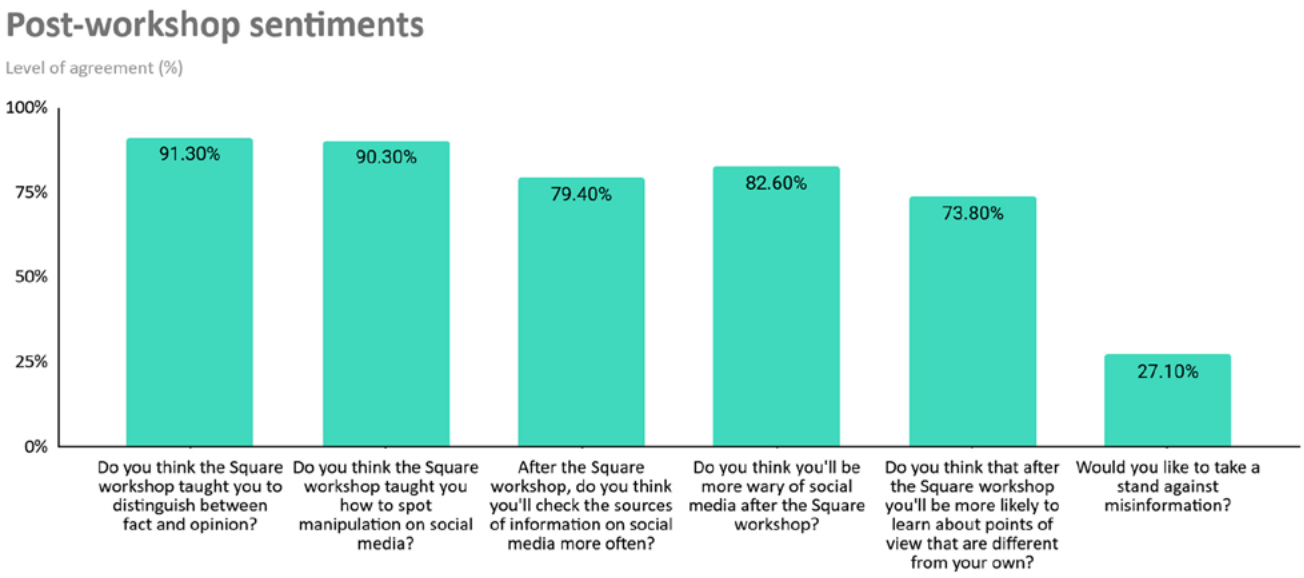
While some individual items showed minor increases in correct responses and decreases in uncertainty, these changes were inconsistent and did not reach statistical significance (N = 29 to 153 at post-test).[44] The average correct response rate rose slightly from 38.3% to 43.1%, but this remains well below a strong performance threshold, indicating that participants still struggle to critically assess digital content in real-world contexts. Similarly, while uncertainty slightly decreased overall, it remained high for many items, suggesting that confidence in evaluating online misinformation did not substantially improve after the workshop.

[44] The list of items presented here focuses on pre-post comparisons, as only one item includes follow-up data. For most items, the available data is either pre-test only, pre-test and post-test, or pre-test and fo low-up. Additionally, the sample size for follow-up testing was much smaller, limiting the ability to draw direct comparisons over time.

Correct response	Visual	Pre-test	Post-test	Pre-test	Post-test
		Correct		*Uncertain*	
TRUE	**Item 1**	66.30%	54.10%	24.90%	32.40%
FALSE	**Item 2**	41.90%	40.90%	38.70%	38.90%
FALSE	**Item 3**	18%	22.90%	57.70%	52.10%
TRUE	**Item 4**	51.10%	48.60%	37.80%	36.80%
TRUE	**Item 5**	12.20%	14.30%	56.10%	53.60%
FALSE	**Item 6**	68.30%	89.30%	29.30%	10.70%
TRUE	**Item 7**	35%	46.40%	47.50%	32.10%
FALSE	**Item 8**	48.80%	71.40%	31.70%	25%
TRUE	**Item 9**	26.80%	28.60%	46.30%	46.40%
FALSE	**Item 10**	14.60%	14.30%	65.90%	53.60%
	Average (all items)	**38.30%**	**43.08%**	**43.59%**	**38.16%**
	Avg. true	38.28%	38.40%	42.52%	40.26%
	Avg. false	38.32%	47.76%	44.66%	36.06%

From intention to action: sustained behavioural changes

Participants were asked on their reflections on the workshop, offering insight into how they perceived its impact on their media literacy skills and online behaviours (N = 183 to 229 at post-test). The responses suggest that the training was effective in improving awareness and critical thinking, particularly in their perceived ability to distinguishing fact from opinion and recognising manipulation on social media. Encouragingly, these self-reported gains align with the actual improvements observed in participants' ability to differentiate fact from opinion. Many also reported being more cautious about online content and more likely to check sources, indicating a shift towards more critical engagement with information.



Participants not only expressed a commitment to engaging more critically with online content after the workshop but also followed through on these intentions months later. The proportion of individuals who reported checking sources more frequently and seeking out diverse perspectives remained stable from post-workshop (N = 184) to follow-up (N = 61), indicating that these behaviours were successfully maintained over time. The most notable shift was a rise in scepticism toward news on social media, with more participants at follow-up expressing greater caution than they initially anticipated. These findings suggest that the workshop had a lasting impact on participants' digital habits, fostering critical engagement and greater awareness of misinformation risks. However, further research is needed to determine whether these self-reported behaviours lead to improved accuracy in assessing online content.

Conclusion - Future directions: Bridging the media literacy gap

This report highlights the urgent need for targeted, evidence-based interventions in media literacy for disadvantaged youth in France.

Across multiple metrics, both middle schoolers and young adults demonstrated gaps in assessing information veracity, with high levels of uncertainty persisting in factual and digital knowledge questions.

Additionally, institutional distrust further exacerbates conspiratorial thinking, a pattern clearly reflected in the data. Confidence in institutions drops sharply from 24.6 percent in middle schoolers to just 9.5 percent in young adults. This distrust correlates with a stronger belief in conspiracy narratives. As institutional scepticism deepens, so too does the reluctance to engage with credible sources, reinforcing a cycle where mistrust and conspiratorial beliefs fuel one another. Given that middle schoolers still exhibit relatively higher trust levels, this stage represents a pivotal window of opportunity for targeted initiatives that can foster lasting institutional confidence before further decline sets in.

Motivated reasoning further complicates engagement with factual information. Despite direct exposure to media literacy training, a substantial proportion of participants continued to prioritise conviction over adaptability. This presents a fundamental challenge for media literacy interventions, as knowledge and skills can be taught, but the willingness to apply critical reasoning in the face of personal biases is much harder to instil. Square's workshops led to meaningful behavioural shifts, with participants reporting increased source-checking, and engagement with diverse perspectives.

While these gains were more pronounced in some areas than others, the findings highlight both the potential of short-term interventions and the challenges of ensuring long-term retention. The variability in impact reflects the complexity of fostering lasting change, particularly within a cohort facing deep-rooted institutional distrust and structural barriers to engagement.

One of the greatest challenges in assessing impact was high attrition across testing points, which is typical for hard-to-reach populations and was particularly pronounced in this study. Distrust-driven survey engagement, participant drop-off, and difficulties in matching responses significantly reduced the number of complete cases available for longitudinal analysis. These disparities in sample sizes across time points, including between Q1 and Q2 for some variables, limited the ability to detect statistically significant effects, even when trends suggested meaningful shifts.

The study also evolved over time, with adaptations to scales and metrics to better reflect local contexts and cognitive accessibility. While these adjustments improved relevance and accuracy, they complicated direct comparisons across testing waves. This means that instances of non-significance should be interpreted with caution, rather than as definitive evidence of no impact. Despite these challenges, follow-up assessments revealed some retention of key media literacy concepts. Overall, follow-up surveys and longitudinal studies are critically needed in this space to enhance current understanding of media literacy and its long-term impact.

Despite these limitations, the report points to meaningful progress in several areas:

- Participants translated intention into action, maintaining their commitment to critical engagement with online content for up to three months.
- Participants reported feeling more engaged in political discussions and more confident in expressing their opinions post-workshop. The strongest gains were in political self-expression, with more participants believing they had opinions worth listening to.
- Participants demonstrated clear improvements in defining key media literacy concepts, with gains in understanding confirmation bias, and algorithmic filtering.
- After the workshop, most participants reported checking sources more frequently, and sought out diverse perspectives, suggesting an increased awareness of verification practices.
- A shift in perception of social media's role was observed, with more participants recognising its commercial motives beyond being just a communication tool.
- Young adults showed improvements in their factual knowledge, with many gains sustained at follow-up.
- Confidence in judging the accuracy of information increased, with fewer «I don't know» responses on several misinformation-related items.

This report reinforces that simply raising awareness about misinformation is not enough. Young people need cognitive tools, confidence, and motivation to actively engage in critical evaluation, source verification, and truth-seeking behaviours. The persistence of high uncertainty, reliance on surface-level verification, and low institutional trust signals deep-rooted vulnerabilities that cannot be addressed through short-term interventions alone. Instead of allowing the often-overlooked youth of France to fall deeper into distrust, conspiratorial thinking, and digital disengagement, we must empower them to think critically, assess information with confidence, and take an active role in society. Media literacy must be seen as an ongoing, structural need, requiring long-term investment, reinforcement, and adaptation to the evolving digital landscape. If left unaddressed, these gaps will only widen, exacerbating social divisions, political disengagement, and misinformation vulnerabilities. However, with targeted interventions, the marginalised youth of France can be equipped not just to navigate the digital world, but to help shape it for the better.

Bibliographic appendix

[1] *CEVIPOF- 2024 Barometer of Political Trust*
[2] *The scale used comes from the International Civic and Citizenship Education Study international survey Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010).*
[3] *Enquête "Ecole et citoyenneté", CNESCO (2018).*
[4] *Pfundmair, M., Wood, N.R., Hales, A. & Wesselmann, E.D. (2024) How social exclusion makes radicalism flourish: A review of empirical evidence. Journal of Social Issues, 80, 341–359. https://doi.org/10.1111/josi.12520*
[5] *See Weir, 2023*
[6] *See Kenzo Nera, Complotisme et quête identitaire, PUF septembre 2023*
[7] *See Cheng & Chen, 2020*
[8] *See Xiao et al., 2021*
[9] *See Warner-Søderholm et al., 2018*
[10] *Stanovich K. E., West R. F. (2007) Natural myside bias is independent of cognitive ability. Thinking & Reasoning, 13(3), 225–247)*
[11] *i.e. bullshit, Newton, C., Feeney, J., Pennycook, G. (2024). On the disposition to think analytically: Four distinct intuitive-analytic thinking styles. Personality and Social Psychology Bulletin, 50(6), 906-923.)*
[12] *Costello, T. H., Pennycook, G., Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. Science, 385(6714), eadq1814.*
[13] *Stanovich K., Toplak M. (2023) Actively Open-Minded Thinking and Its Measurement, J.Intell., 11(2), 27*
[14] *See Haran et al., 2013; De Keersmaecker & Roets, 2017; and Pennycook et al., 2023*
[15] *See Kociv & Füchslin, 2018*
[16] *Jolley, D., Douglas, K. M., Skipper, Y., Thomas, E., & Cookson, D. (2021). Adolescent Conspiracy Beliefs Questionnaire (ACBQ) [Database record]. APA PsycTests. https://doi.org/10.1037/t82252-000*
[17] *See Jolley et al., 2021*
[18] *See Larsen et al., 2020*
[19] *See Kociv & Füchslin, 2018*
[21] *Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., Mitchell, R. C., Presser, S., Ruud, P. A., Smith, V. K., Moody, W. R., Green, M. C., & Conaway, M. (2002). The impact of «no opinion» response options on data quality: Non-attitude reduction or an invitation to satisfice? Public Opinion Quarterly, 66(3), 371–403. https://doi.org/10.1086/341394*
[22] *KGuay, B., Berinsky, A. J., Pennycook, G., & interventions work. Nature Human Behaviour, 7(8), 1231-1233.*
[23] *Pfänder, J., & Altay, S. (2025) Spotting False News and Doubting True News: A Meta-Analysis of News Judgements. Nature Human Behaviour.*
[24] *Pennycook, G., Binnendyk, J., Newton, C., & Rand, D. G. (2021). A practical guide to doing behavioural research on fake news and misinformation. Collabra: Psychology, 7(1), 25293.*
[25] *Pfänder, J., & Altay, S. (2025) Spotting False News and Doubting True News: A Meta-Analysis of News Judgements. Nature Human Behaviour.*
[26] *Pennycook, G., Binnendyk, J., Newton, C., & Rand, D. G. (2021). A practical guide to doing behavioural research on fake news and misinformation. Collabra: Psychology, 7(1), 25293.*
[27] *Badrinathan, S. (2021). Educative interventions to combat misinformation: Evidence from a field experiment in India. American Political Science Review, 115(4), 1325-1341.*
[28] *Lühring, J., Shetty, A., Koschmieder, C., Garcia, D., Waldherr, A., & Metzler, H. (2024). Emotions in misinformation studies: Distinguishing affective state from emotional response and misinformation recognition from acceptance.*
[32] *Rosenberg, M. (1965). Society and the adolescent self-image. Princeton, NJ: Princeton University Press.*
[33] *CEVIPOF - Baromètre de la confiance politique / Vague 12 – Février 2021*
[37] *Imhoff, R., Bertlich, T., & Frenken, M. (2022). Tearing apart the 'evil' twins: A general conspiracy mentality is not the same as specific conspiracy beliefs.*
[43] *See O'Mahony e al., 2023*

Annex 1 – Case-by-case study of middle schoolers' discernment skills

by Manon Berriche

In 2024, Square decided to further complement the assessment of perceived media literacy skills by evaluating the reactions and abilities of middle-schoolers' when facing real-world informational content on social media. Using an online software tool (Wooclap), 750 students aged 11-15 were presented with cases from TikTok, Meta and Instagram. Through a computer they were tasked to answer, according to them, whether the screenshot in front of them was 100% true, likely true, likely false, 100% or whether they didn't know (N = 222 to 615 depending on items). They were also encouraged to comment and explain their choices through a textual queue, and a small part of them did so. This annex focuses on these qualitative comments from the students (N = 73 to 183 depending on items). Students' open-ended responses enrich the interpretation of quantitative analyses by shedding valuable light on their motivations and the criteria used to evaluate content. They help us to better understand why certain content is accepted or rejected, and to identify which cues they consider reliable and which mislead them.

Among the students' verification and evaluation strategies, we can first mention attention to the source of the content. Some students report having heard of «Le Parisien» or «Le Figaro» and see them as reliable media specializing in the production and dissemination of important news. This reference to established sources helps to anchor information in a recognized and credible context. Next, many students pay attention to linguistic or visual cues. Some students detect markers of uncertainty in the language used, such as the use of expressions like «imagine» or «if this is confirmed», which signal unsettled information. Similarly, particular attention is paid to images: the possibility of editing or the use of artificial intelligence leads some students to doubt the visual authenticity of the content.

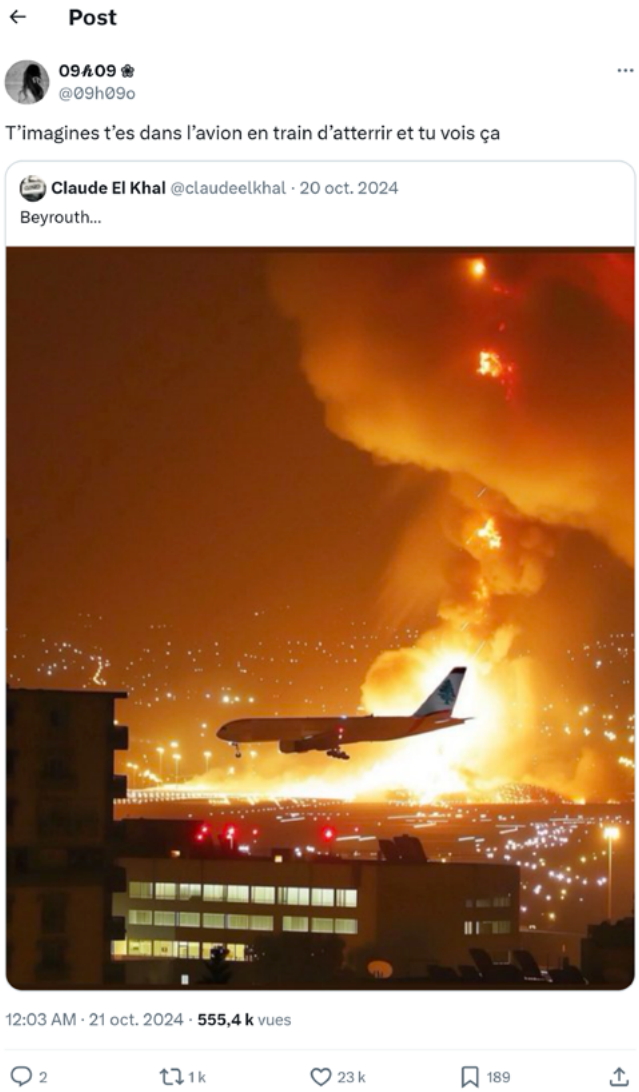
Finally, a significant number of students base their assessment on logical analysis, comparing information with prior knowledge and using common sense. This approach encourages a more nuanced analysis, even if it can sometimes remain superficial if not coupled by a thorough verification of the facts.

On the other hand, what can mislead them includes reliance on the official account without thorough analysis, the belief that if a similar event has already occurred, the information is necessarily true, or overconfidence in the perceived reality of the image. It's interesting to note that, even if a student's answer can be considered wrong from a quantitative point of view, qualitative analysis often reveals a certain reflexivity, as evidenced by the statement: «Anything can happen, but that doesn't mean we shouldn't be suspicious». Conversely, some students give «good answers» whose arguments remain superficial or indicate a difficulty in understanding.

A significant proportion of students are uncertain about informational content, answering «I don't know» rather than making up their minds. This cautious attitude reflects both a lack of verification elements (absence of source, insufficient explicit information) and an inability to identify the clues that would enable them to correctly assess the reliability of a content, such as a reliable source or visual and textual coherence. Example, Beirut Case: «I don't know if it's true or false, there's no source»; «In the image, there's no info to see if it's true or not»; «It could be false as well as true»; «You can't tell if it's true or false»; «There's nothing to prove to me that it's true.»

Case - Plane Beirut

Students judge the image to be credible, based on :
• a partial knowledge of current events and an impression of déjà-vu, referring in particular to the geopolitical context in Lebanon (e.g. «It must be true because the country is currently at war»; «Because I know that there is war in Lebanon, so for me it's true»), or more rarely to recent fires in Los Angeles (e.g. «because I saw it on the news in Los Angeles»),
• an absence of obvious signs of manipulation (e.g. «because there's no Photoshop or green background»; «the image doesn't look faked»),
• the plausibility of a plane crash or fire (e.g. «because planes can explode»; «fires in the area are very likely»; «because it can happen»; «sometimes there are fires in certain countries»; «planes can crash»),



• the perceived realism of the image (e.g. «Duh the image looks like reality»; «I think it's true because the image is very realistic»; «the fire looks real»). Here, evaluation is based on a perceptual analysis of the image, rather than on the source and content of the information itself.
Students express hesitation, wondering about the reliability of the source and/or the possibility of image manipulation, but without any explicit methodological tools for verifying the information (e.g. «There's nothing to prove to me that it's true»; «I don't know if it's true or false, there's no reliable source»; «It's easy with technology these days to create an image thanks to AI»; «We don't know if the source is reliable or not and the image is probably generated by Artificial Intelligence»; «The photomontage exists so I don't know if it's true»).
Some students display a general skepticism, arguing that «most things on social media are lies» and that «not everything you see on social media should be believed». A few comments suggest that it's a montage, or that the image looks fake, without always backing up their answers («it looks like a montage», «in the back wing, you can see that it's been added», «it looks like a cartoon»).
Other students demonstrate more critical thinking by pointing out the absence of reliable sources (e.g. «If it had been Le Parisien, maybe, but here...»; «The account that posted this photo is not an official TV account.») or by pointing out linguistic clues that invite doubt (e.g. «Because it says 'imagine'»). Finally, many students draw on their knowledge of how airplanes and airports work and consider the image presented to be unrealistic, insisting that the plane is not destroyed in the photo or that, under real conditions, a plane would not fly (so low) or land in a residential neighborhood (e.g. «Because if there was a real explosion, the runway would be blocked.»; «Planes don't fly so low near buildings.»; «A plane doesn't land in a neighborhood. Or it doesn't fly that low»; «Because a plane can't be that close to the ground except to land»; «Duh a plane that flies very low it's not too logical we only see that in the movies»).

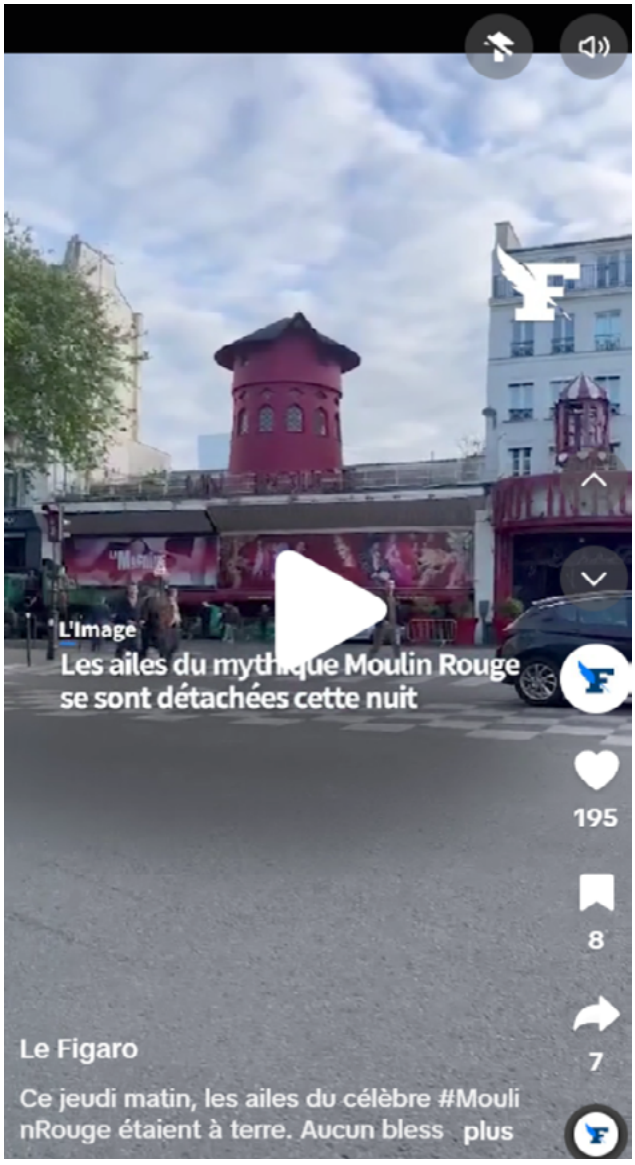
Case - Grande Motte Synagogue

A majority of students consider the information to be 100% true or probably true, based mainly on the fact that the account is certified. The recurring argument is based on the idea that certification guarantees the veracity of the content: «The account is certified, so it's true». Some students justify their adherence to the information simply because of the realism of the photo, without taking into account the content of the tweet (e.g. «because it's a real photo; «the image looks like a real one»; «it looks real»).

A significant proportion of students seem perplexed by the content. Several said they had «understood nothing». Others mention the absence of any trace of fire in the photo and consider that they cannot comment on the credibility of the tweet (e.g. «I can't see the fire, so I can't say whether it's false or true»).

A minority reject the assertion by identifying linguistic clues in the tweet that invite doubt («It says 'imagine'»; «This person wrote 'if confirmed' which means it's not confirmed»). Some also point to a lack of additional information, or suspect potential manipulation, prompting them to put the veracity of the information into perspective (e.g., «There's no information, it may have been altered», «It may have been edited»).

Case - Moulin Rouge



The majority of students accept information based on the authority of the source. For example, they say: «It's 100% true because it's relayed by Le Figaro, which is reliable» or «I don't think Le Figaro would say things that aren't true». Several students also mention having already heard about this information, claiming to have seen it a few months ago on the 8 o'clock news (e.g. «I saw a few months ago on the news that it was true»; «it's true, newspapers talked about it»). In addition, some rely on visual analysis of the image, noting the absence of wings on the mill - comments such as «you can't see the wings anymore» or «there are no wings» illustrate this point. Finally, a few students justify their position by arguing that the incident seems plausible to them, noting for example that «it's possible on a windy night» or that «it could be true, because there's a lot of wind». These latter answers are based more on immediate intuition than on in-depth reflection.

Some students adopt a cautious stance, expressing their lack of knowledge and information, for example: «I don't know, I don't know this mill»; «I don't have enough information to answer this question»; «I've never heard this story».

Finally, some students reject information despite its veracity, using arguments that they present as critical, but which often remain superficial. For example, some are content with a literal reading, stating that «windmills don't have wings, they're not birds» or that «a windmill doesn't fly». Other arguments are based on a simplistic application of common sense or superficial knowledge (e.g. «it's wrong because a monument like this should be safe and secure»; «I doubt that the builders of the Moulin Rouge didn't manage to make solid wings»; «the mill is quite solid and it can't fall down because even if it's old it needs to be renovated»). Finally, some students feel that the information is false because there aren't enough likes (e.g. «not many people liked it»; «it's false, because if it were true, there would have been more likes and comments»).

Case - El Mordjene

The majority of students immediately accept the veracity of the information, based on elements of authority and media recognition.
For example, they say: «It's true because it's broadcast by the newspaper Le Parisien»; «Because Le Parisiens is a channel that explains current events and what happens in a day»; «For me it's true because we can see that the account is certified». Many also mentioned that they had already heard of this information and claimed to know that the export of dairy products to Europe is forbidden, suggesting that the content had been widely circulated in their environment (e.g. «It's true because I saw this info on the internet and he had explained that the spread had milk and it couldn't come to Europe»). However, although most students correctly evaluate the information on a factual level, their reasoning is sometimes based on fallacious or simplistic arguments. Indeed, rather than supporting their statements with in-depth analysis, some students simply criticize France's decision, adding value judgments: «They banned it because they're racist and the woman has a veil», «Everyone knows it's true, because they don't want a copy of Nutella», «It's bankrupting Nutella, they're racist», «France is a country jealous of Algeria's popularity», or «It's true because Europeans are mad about El Morgen». These answers, while factually correct, show that some students are content with emotional and superficial argument, rather than rigorous fact-checking.

The level of doubt about this content is particularly low. Comments expressing uncertainty are rare and are often limited to formulas such as «I don't know», which does not allow for more in-depth interpretations.

Finally, some students reject the information despite its veracity, basing their responses on personal experience rather than an overall analysis of the situation: «I have some at home»; «My aunt bought some»; «It's not true because my cousin's cousin bought el mordiene in France»; «Because you can still buy it». Other responses illustrate a lack of understanding of the regulatory mechanisms governing food safety in France («Because they can't do anything, the authorities don't have the right to ban, it's not true»; «Why would they do that?»; «You can buy whatever food you want.»).

Case - Anne Hidalgo

Some students immediately accept the veracity of information, even if it's false, based on partial knowledge or having heard of it before. For example, they say: «I think it's true because Anne Hidalgo is the mayor of Paris»; «I'm pretty sure it was on TV». Simply recognizing the name of a politician they've already heard of seems to be enough for them to assess the information as credible. Many students don't seem to know Le Gorafi and so don't identify the parodic intention behind the content. Some even identify the source as a news channel: «I think it's a newspaper»; «Because Le Gorafi is a channel specializing in news»). Finally, a few students judge the information to be plausible, using superficial reasoning (e.g. «it could be true because there are a lot of pedestrians in Paris»; «because there could be accidents»; «because there are too many pedestrians in the streets of Paris today»).

Some students adopt a cautious stance, indicating that they don't have enough information to make a judgement (e.g. «I don't know because I don't have enough information on that, nor am I sufficiently educated on the subject»). These responses reflect uncertainty and a lack of knowledge.

Finally, some students explicitly reject the information, appealing to their logic and common sense. They put forward arguments such as: «It's technically impossible»; «It can't limit pedestrians' time, pedestrians may be very old or have a broken leg». Some students analyze the content in greater depth, assessing the credibility of the source: «It's an unknown site and the views/likes are few»; «The account isn't certified», and are aware of Gorafi's parodic intent («It's a counterfeit Figaro channel»).

Cases - Meta/Israeli

Some students immediately accept the veracity of information by relying on references of authority. For example, they affirm that the information is true by pointing out that the account is certified (e.g. «It's a certified account», «certified account», «because the account is certified»).

Some students invoke the moderation rules of social media to justify their answers (e.g. «because we're not allowed to publish racist things, it's possible»). Many answers, although quantitatively correct, indicate that the students have not fully understood the content: many find credible the fact that the Irish women did not shake hands with the Israelis, based in particular on their knowledge of the conflict between Israel and Palestine, but do not assess the credibility of Meta's decision to have closed the Irish basketball team's account (e. g. «In my opinion it should be closed», «In my opinion it must be true that women basketball players don't want to shake hands with Israeli women, especially when we know what's happening to Palestinians because of them»; «because maybe women basketball players support Palestine»; «given the geopolitical situation in the Middle East, many people defend the oppressed»).

These responses suggest that students have read the content too quickly. Finally, a few students use the comments to give their opinion on the Israeli-Palestinian conflict rather than to back up their answers with arguments, suggesting that they have evaluated the content correctly because they confirm their beliefs but not critically (e.g. «They did the right thing»; «FREE PALESTINE»; «For me it's true because I don't like Israelis!!!; «it's true and I agree Israel doesn't deserve to be shaken hands with»).

A second category includes students who express their lack of information to make a definitive statement («I don't know», «I have nothing to say», «I haven't seen it») or the absence of a reliable source («It's not from the official Meta account, so we don't know»). Many say they have no interest in basketball and know nothing about this team («I don't watch basketball»; «I've never heard of this team»).

Finally, a minority explicitly rejects the information, invoking the principle of fair play and relying on the idea of an obligation to shake hands in sport (e.g. «Because all players in a sport are obliged to shake hands with their opponents»; «Because it's respectful to shake hands with other players»). What's more, some people seem to misunderstand Meta's moderation rules, or at least find them exaggerated, asserting, for example, «We're not going to close an account because of this»; «Just because they didn't shake hands doesn't mean their account will be deleted»; «Just because you don't shake hands with Israelis doesn't mean an Instagram account will be deleted»; «This is not a reason to delete an account.»

Case - Pyramids Alaska

Very few students judge content to be credible based on popularity indicators such as a high number of likes and views (e.g. «There are a lot of views»).
Several students are cautious in expressing the idea that, if the reported event were real, it would necessarily have generated international media coverage and word of mouth (e.g. «If it had been true, it would be worldwide»; «I've never seen such a thing»; «I've never heard of it»).

Most comments were made by students who rated the content as «100% false» or «probably false». Some students mobilize elements of general knowledge and relying on principles of plausibility or feasibility (e.g. «Because the pyramids were in the desert where it's hot, not in Alaska where it's cold.»; «Alaska is not in the desert»; «You need sand to build pyramids»; «There are no pyramids in America, let alone underwater»; «This image is false because the water is too cold to go underwater and make a pyramid»; «It's impossible»; «It may be impossible»).

These arguments reflect a desire to assess the material feasibility of the phenomenon described and show an attempt to mobilize historical and geographical knowledge.

However, some assertions are approximate, and the analysis often remains superficial, suggesting partial understanding or preconceived ideas.

Several comments express skepticism about the source of the information, notably because of the lack of certification or the context of dissemination: «The information on TikTok is mostly false.»; «It's probably false because the account isn't certified.» This type of reasoning shows a certain awareness of the criteria for assessing the credibility of a source, even if the notion of certification is not always an absolute guarantee of reliability. A few comments suggest an awareness of the dynamics of virality on social networks: «It's a trend». These students seem to recognize that certain information can be disseminated simply to attract attention, without being based on fact. However, none of the students noted the sensationalist tone of the content (presence of numerous capital letters and sensationalist expressions such as «shocking revelations»).

Some comments express value judgments or strong opinions without any real argumentation: «The Americans are too stupid, so I don't believe it». This type of response shows the impact of stereotypes and pre-existing beliefs on the evaluation of information.

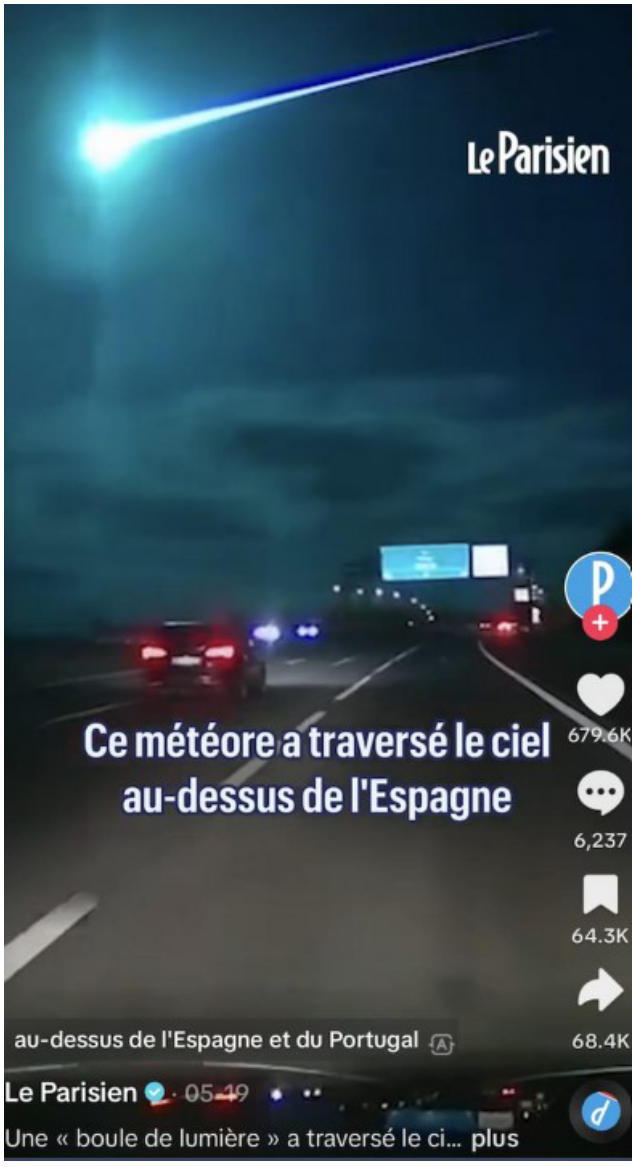
Case - Marine Le Pen / homelessness

The information is considered true because Marine Le Pen is perceived as «racist» and because she is said to have expressed positions aimed at excluding non-French people, which, according to them, would confirm the veracity of the message (e.g. «It's true because Marine Le Pen wants people who aren't 100 percent French to leave the country»). Some students express personal or ideological criticism, as in «Marine Le Pen is a racist, so it's true». This type of response illustrates a tendency to associate a clear-cut opinion with a generalization based on a moral or political judgment, without arguments backed up by verifiable facts.

A significant proportion of students remain undecided, positioning themselves in a state of uncertainty when faced with information.
Some say that they don't often watch the news, or that they don't have enough evidence to make up their minds: «To tell you the truth, I don't know at all, because I don't often watch the news». Others express hesitation, noting that, despite being broadcast on the networks, certain elements give rise to doubts («This information may be true, but at the same time false, because there are certainly a lot of views, but that doesn't mean it's true»). These answers show a certain lucidity on the part of the students as to the limits of their knowledge, even if they also reflect a difficulty in confronting their opinion with the information available.

A significant proportion of students reject the information, using critical arguments. They point to the lack of media coverage («nobody mentioned it in the news») as an indication of non-veracity. Some express their rejection by invoking value judgments (e.g., calling it unfair or racist), or by appealing to their logic or common sense («It's already shit in France, so they're going to give money to those who denounce the undocumented»; «People don't have that much to do»).

Case - Meteor Spain



Students who accept the veracity of information rely heavily on markers of media authority. They mention the dissemination of the information by recognized media outlets, and emphasize the certification of the account as a guarantee of credibility: «Le Parisien is a 100% truthful newspaper»; «The 'Le Parisien' account is certified»). Some also mention their recollection of having seen it on several media (television, TikTok): «It was on social media» , «True, I've seen the video», «It was on TF1, it was said in the news».

Some students remain undecided and express their uncertainty. They often content themselves with statements such as «I don't know», or express doubts about the reality of the phenomenon, for example concerning the existence or visibility of meteors («I don't know if meteors exist and if you can see them in the sky»). These responses reflect caution, which can be explained by a lack of knowledge or by the complexity of the subject.

Finally, some students explicitly reject the information, citing several arguments: (1) the possibility of editing (e.g. «Photoshop», «You can do anything with AI», «there are filters, you can see it»); (2) the lack of credibility of content circulating on social media (e.g. On TikTok, most people lie for the buzz). These responses testify to a distrust of social media content, associated with manipulative practices (editing, artificial intelligence).

Annex 2 –

Summary of Quentin Daviot and Simon Briole's analysis of data collected in 2023

Tableau 7 : effets généraux et hétérogènes de l'atelier Square sur les dimensions d'intérêt

		Effets hétérogènes selon caractéristiques, perceptions et connaissances mesurées avant le début de l'atelier								
Dimensions d'intérêt	Général	Complotisme fort	Confiance en soi politique faible	Ouverture d'esprit faible	Connaissances testées faibles	Estime de soi faible	Confiance dans les institutions faible	Satisfaction de vie faible	Faible capacité de distinction fait/opinion	Homme
Complotisme	-0,38***	-1,08***	-0,27**	-0,11	-0,19	-0,35**	-0,42***	-0,28**	-0,13	-0,44***
Ouverture d'esprit	-0,19	-0,26	-0,24	1,29**	-0,19	-0,30	-0,2	-0,17	0,31	-0,17
Confiance en soi politique	0,32***	0,37***	1,13***	0,37***	0,31**	0,32**	0,19*	0,28***	0,27**	0,33***
Compétences déclarées	0,49***	0,35*	0,79***	0,63***	0,58***	0,62***	0,56***	0,39***	0,58***	0,37*
Connaissances testées	0,76***	0,90***	1,06***	1,06***	2,60***	0,71***	0,90***	0,73***	1,00***	0,86***
Distinction fait/opinion	0,36***	0,26*	0,37***	0,52***	0,54**	0,36**	0,45***	0,46***	1,00***	0,39**
Nombre d'observation moyen	122	68	60	48	52	62	60	78	62	53

Note : Le nombre d'observations utilisées dans l'analyse peut varier d'une dimension à l'autre selon le profil de réponse des répondant·es (par exemple, certain·es n'ont pas répondu à toutes les questions du questionnaire). Nous reportons donc ici le nombre moyen d'observations utilisées dans l'analyse des effets.
* : significatif à 10% , ** : significatif à 5% , *** : significatif à 1%.
Lecture : Selon la méthodologie avant/après utilisée pour cette évaluation, les ateliers Square permettent en moyenne de diminuer le complotisme de 38% d'un écart-type (au seuil de significativité de 1%). Pour les jeunes qui montraient un niveau de complotisme élevé avant le démarrage de l'atelier (c'est-à-dire avec un niveau de complotisme supérieur à la médiane), l'atelier Square permet en moyenne de diminuer le complotisme de 1,08 écart-type (au seuil de significativité de 1%).