

Algorithmic Transparency, User Control, and Adolescent Self-Efficacy: A Review and Research Agenda (2020–2025)

Wenting Su^{1,*}

¹School of Foreign Language,
Shenyang Jianzhu University,
Shenyang, 110168, China

*Corresponding author: wsu0202@gmail.com

Abstract:

Personalized recommendation algorithms structure what adolescents see and when they see it, potentially shaping self-regulatory outcomes. This review synthesizes recent empirical work (2020–2025) on how algorithmic transparency and user controls relate to adolescents' task specific self-efficacy and metacognitive processes (planning, monitoring, evaluation). Evidence from large surveys links higher general self-efficacy to fewer emotional symptoms even when negative social media experiences are considered. Mixed methods and experimental interface studies indicate that explanations (e.g., “why am I seeing this?”) and steerable controls (e.g., topic filters, reset/diversify) are associated with greater perceived transparency, trust, and reflective engagement, while low transparency, engagement optimized ranking is associated with diminished perceived control. Converging across educational and open platform contexts, these patterns suggest that clarity and controllability may help preserve adolescents' efficacy beliefs and support metacognitive regulation. However, most studies rely on cross-sectional designs or proximal indicators (e.g., trust, connectedness) rather than direct metacognition/efficacy measures, and field experiments on live platforms remain scarce. This review outline design implications (plain language explanations; visible impact views; low friction steering) and educational directions (algorithmic literacy instruction), and propose a research agenda emphasizing preregistered field studies, validated instruments for efficacy/metacognition, and transparent analytic reporting. Together, these steps can link algorithmic design choices to measurable benefits for adolescent users across cultural and platform contexts.

Keywords: Recommendation algorithms; adolescents; self-efficacy; metacognition

1. Introduction

Recommendation algorithms now curate a substantial share of adolescents' media environments, ranking and personalizing content to maximize engagement. For young users, this architecture can alter perceived control over information flows and introduce uncertainty about how behaviors translate into exposure. Within developmental and educational psychology, two constructs are particularly informative for understanding these effects: self-efficacy (adolescents' belief that they can organize and execute actions to reach goals in digital settings) and metacognition (the knowledge and regulation of one's cognition, including planning, monitoring, and evaluation during use).

Contemporary studies highlight both risk and opportunity. Largescale survey evidence associates higher general self-efficacy with fewer emotional symptoms in youth samples controlling for social media exposure and negative experiences [1]. Mixed methods research with youth juries and focus groups documents that opacity raises concerns about privacy, safety, trust, and perceived control, and that adolescents favor clearer explanations and meaningful user controls [2]. Interface experiments and adolescent centered design programs show that "why" explanations, "what if" impact views, and steerable controls can increase perceived transparency and invite reflective engagement—proximal states relevant to planning and monitoring [3-5]. Related diary and survey studies on perceived algorithm responsiveness suggests links to social connectedness that may cooccur with reflective evaluation, though metacognition itself is seldom measured directly [6]. At the same time, concerns about reward-oriented design and compulsive checking justify attention to emotion regulation and attention control [7,8].

Despite this progress, several gaps remain. Causal identification is limited because many results derive from cross sectional or short-term designs—typically single-session interface experiments or diary studies lasting only several days to a few weeks—and outcomes often rely on proximal cognitions (e.g., trust, responsiveness) instead of validated metacognition or task specific efficacy scales. Evidence from live social platforms, where ranking is rapid and stakes are high, remains sparse compared with educational prototypes. Cross cultural and cross platform comparisons are also underdeveloped, complicating generalization.

This review consolidates empirical findings on recommendation algorithms' relationships to adolescents' self-efficacy and metacognition, articulates the mechanisms by which transparency and steerability may influence perceived control and reflective engagement, and develops design, educational, and policy implications. This review

concludes with a research agenda calling for preregistered field experiments, validated instruments for metacognition and task specific efficacy, multiwave designs to assess durability, and transparent reporting of analytic choices and model sensitivity, to link design decisions to measurable benefits for adolescent users across cultural and platform contexts [9,10].

2. Conceptual Framework and Key Constructs

2.1 Recommendation Algorithm Mechanisms

In this review, recommendation algorithms are machine-learning systems that assemble personalized, ranked feeds by estimating attention, click-through, and dwell time, then tuning subsequent exposure to maximize engagement. For adolescents, such systems often operate with limited transparency and thereby shape both what is shown and when, altering the felt locus of control over the media environment. When designers add clear explanations or user-facing controls, young users report improved comprehension of how recommendations are generated and greater trust in the system's fairness and reliability—suggesting that transparency scaffolds more critical reasoning about algorithmic influence [2,4].

2.2 Self-Efficacy

Self-efficacy is treated here as adolescents' belief that they can organize and carry out actions to achieve desired outcomes in digital settings (e.g., staying on task, moderating use, learning effectively). Empirically, general self-efficacy functions as a coping resource linked to fewer emotional symptoms [1]. At the same time, engagement-maximizing loops can monopolize attention and narrow perceived options, eroding perceived control and, in turn, undermining efficacy, especially when compulsive, reward-driven checking characteristic of behavioral addiction and emotional dysregulation emerge [8].

2.3 Metacognition

Metacognition encompasses knowledge about one's cognition and the regulation of it—planning, monitoring, and evaluation during media use and learning. In school-age populations, metacognitive development can be intentionally supported and assessed [9]. Within algorithmic environments, algorithmic literacy or awareness captures users' ability to notice personalization, anticipate its effects, and adapt strategies to mitigate personalization effects—e.g., diversifying sources, adjusting follow/like behavior, or using 'not interested' and topic filters (e.g., diversifying

inputs, ignoring cues, using controls). Design work shows that surfacing explanations and providing granular learner controls can facilitate this reflective monitoring and strategy adjustment [3,5].

2.4 Mechanisms Linking Algorithms, Metacognition, and Self-Efficacy

Conceptually, recommendation algorithms structure the stimulus field and reward contingencies adolescents face. Low transparency coupled with strong engagement optimization can place heavy demands on metacognitive monitoring while reducing perceived control, a core condition for maintaining self-efficacy. Conversely, promoting algorithmic literacy and providing meaningful controls and explanations can support the full cycle of metacognitive regulation and help preserve task-specific self-efficacy in digital contexts [6,7].

3. Impacts of Recommendation Algorithms on Adolescents' Self-Efficacy

3.1 Theoretical Pathways and Hypotheses

Building on Section 2, this chapter examines how recommendation algorithms shape adolescents' task-specific self-efficacy in digital contexts.

3.2 Evidence from Large-scale Survey Research

Study 1: Using a national adolescent survey from Norway (Ungdata 2021; $N = 139,841$), the authors estimated multivariate linear regressions predicting depressive symptoms from daily time on social media, four negative social-media experiences (online bullying; family conflict related to social media; too little sleep due to social media; pressure for followers/likes), and general self-efficacy (5-item General Self-Efficacy Scale). In multivariate linear regressions, higher general self-efficacy was associated with lower depressive symptoms (Model 2 $\beta = -0.29$, $p < 0.001$), explaining an additional 7.9% of variance after social-media exposure variables; effects remained after sociodemographic adjustment (Model 3 $\beta = -0.26$, $p < 0.001$). Negative experiences showed medium positive associations (e.g., online bullying $\beta \approx 0.19$; pressure for followers/likes $\beta \approx 0.13$, all $p < 0.001$). Gender-stratified models indicated stronger protective associations for girls ($\beta = -0.32$) than for boys ($\beta = -0.22$), both $p < 0.001$. Limitations include cross-sectional design, self-report measures, and unobserved confounding (e.g., prior mental-health history, family SES, offline stressors). This pattern suggests that bolstering self-efficacy may buffer symptoms even when negative platform experiences occur

[1].

Study 2: Across two preregistered studies (an online survey and a 14-day daily diary with TikTok users), perceived algorithm responsiveness (i.e., the sense that the feed reflects one's identity and goals) predicted greater social connectedness, whereas perceived algorithm insensitivity related to lower connectedness. Although self-efficacy was not directly measured, the findings implicate perceived responsiveness/control as proximal socio-cognitive mechanisms relevant to adolescents' efficacy in navigating algorithmic feeds [6].

3.3 Evidence from Experimental Interface Interventions

Study 3: In a randomized between-subjects lab study with adolescents ($N = 76$; 12–19 years), participants interacted with an e-learning recommendation tool under three conditions: (1) no explanation/controls (NONE), (2) a basic control panel to steer exercise difficulty (e.g., selecting easier/harder tasks) (CONTROL), and (3) control + impact visualization showing how steering changed the recommended set (CONTROL+IMPACT), compared to a no-control baseline (NONE). One-way ANOVAs indicated that both explanation variants significantly increased perceived transparency and trust relative to NONE, with the CONTROL+IMPACT condition yielding the highest ratings ($p < 0.05$). Although self-efficacy was not directly assessed, the combination of user controls and visible causal impact plausibly enhances perceived agency—an antecedent to task-specific self-efficacy in digital learning [4].

Study 4: Using mixed methods (youth juries/focus groups plus a survey; total $N = 260$ young people), the study documented benefits of personalization alongside concerns about privacy, safety, and opaque decision-making. Participants favored explainability and meaningful user control, reinforcing the design hypothesis that transparency and controls can support adolescents' sense of efficacy when engaging with algorithmic systems [2].

3.4 Complementary Learning-Analytics Evidence on Efficacy Modeling

Study 5: Secondary analyses of OECD Programme for International Student Assessment (PISA) datasets with cross-national samples of 15-year-olds using machine-learning approaches showed that students' self-efficacy can be predicted from large-scale behavioral and contextual indicators. While not about social media, these results demonstrate that self-efficacy is a measurable construct whose sensitivity to model specification is evidenced by its stable feature importance across algorithms,

its incremental predictive gain over demographics and time-spent, and its acceptable calibration, thereby motivating future work to instrument platform-level efficacy signals in algorithmic environments [10].

3.5 Synthesis, Limitations, and Implications

Across heterogeneous methods, the evidence converges on two points. First, self-efficacy functions as a protective resource in adolescent digital contexts: higher efficacy correlates with fewer symptoms even when exposure and negative experiences are controlled [1]. Second, interface-level transparency and user controls can improve proximal cognitions (trust, perceived transparency, perceived responsiveness) that are theoretically upstream of efficacy beliefs [2,4,6].

However, causal identification remains limited. Most evidence is cross-sectional or relies on indirect indicators (e.g., trust, connectedness) rather than direct efficacy outcomes; rigorous pre-registered field experiments that measure domain-specific self-efficacy (e.g., learning-task efficacy, distraction-control efficacy) are scarce. Measurement choices vary—many studies use global General Self-Efficacy rather than task-specific scales. This paper recommends that future trials combine transparent recommendation controls with validated efficacy scales and behavioral performance metrics, and assess moderators such as gender and baseline algorithm literacy. From a design perspective, adding explainability and steerable controls to adolescent-facing feeds is a low-risk intervention likely to bolster perceived agency and, by extension, self-efficacy [7].

4. Impacts of Recommendation Algorithms on Adolescents' Metacognition

4.1 Theoretical Pathways and Hypotheses

Building on Chapter 2, metacognition is examined as planning, monitoring, and evaluation during algorithm-mediated use. Engagement-oriented ranking can obscure how outcomes arise, which may raise monitoring demands and discourage evaluation. Transparent explanations and user controls are expected to facilitate adolescents' reflective monitoring and strategy adjustment; thus, metacognition should covary negatively with opacity and positively with explainability and steerability [2-5].

4.2 Evidence from Empirical Studies

Study A: Participants comprised adolescents and educators involved in iterative, adolescent-centered design studies. The research employed co-design sessions and

user evaluations of learner controls and visual explanations that clarified why items were recommended and how steering would change results. Analyses combined qualitative feedback with usability and perceived-transparency measures. Findings indicated that adolescents demonstrated positive engagement with 'what-if' impact views that clarified the consequences of steering, while teachers reported that "why" explanations supported lesson planning. The study concluded that steerability and clear feedback support planning and monitoring. Limitations include early-stage prototypes, predominantly qualitative evidence, and the need for confirmatory trials [5].

Study B: The study recruited young people for youth juries and focus groups, supplemented by a survey on experiences with algorithm-mediated platforms. The mixed-methods design captured perceived benefits (discovery, convenience) and concerns (privacy, safety, trust, perceived control). Qualitative analysis and descriptive survey statistics showed that opacity undermined perceived control and trust; these perceptions may weaken adolescents' metacognitive strategies for regulating use, including planning, monitoring, and evaluation. The authors concluded that transparency and safeguards are necessary for young users. Limitations include non-probability sampling and the absence of direct metacognitive scales [2].

Study C: In a randomized laboratory experiment with adolescent users of an AI learning recommender, participants interacted with either a baseline interface or versions that added steering controls and an impact visualization. One-way ANOVAs showed higher perceived transparency and trust under explanation conditions, with the impact visualization yielding the highest ratings. Although metacognition was not measured directly, the outcomes align with reflective monitoring and evaluation supported by explanations and controllability [4].

Study D: Two preregistered studies—an online survey and a two-week daily diary with TikTok users—introduced perceived algorithm responsiveness and insensitivity to model links between personalization and social connectedness. Higher perceived responsiveness was associated with greater connectedness, a state that can indirectly support reflective evaluation and goal-directed monitoring by signaling alignment between user intentions and feed content. While metacognition was not directly measured, the pattern is consistent with reflective evaluation and goal-directed monitoring during use [6].

4.3 Synthesis and Comparison across Studies

Across diverse methodologies, studies converge on the importance of transparency and user control for metacog-

nitive processes. Experimental and design work shows that explanations and steerability are associated with higher perceived transparency and reflective engagement, whereas mixed-methods evidence documents concerns about opacity and perceived loss of control. Differences arise in context and measurement: e-learning prototypes capture proximal outcomes such as perceived transparency and strategy reflection, while open-platform surveys emphasize connectedness rather than direct metacognitive scales. Taken together, these findings suggest that making ranking logic visible and controllable can support planning, monitoring, and evaluation in adolescent use.

4.4 Limitations and Directions for Future Work

Most studies rely on short sessions, self-report outcomes, or early-stage prototypes, and direct metacognitive instruments are rare in social-media settings. Future work should implement pre-registered field experiments on open platforms, include validated metacognitive scales alongside behavioral indicators, and test moderators such as age, baseline algorithmic literacy, and task demands [5,6].

4.5 Practical Implications

Designing adolescent-facing feeds with plain-language explanations, visible impact views, and low-friction controls can help sustain reflective monitoring and strategy evaluation. Instruction in algorithmic literacy—potentially integrated into school curricula—can further reduce monitoring burden and support goal-directed engagement [2-5].

5. Discussion

5.1 Cross-study Synthesis

Findings across Chapters 3 and 4 point to a shared linkage: adolescents' perceived control over transparency, personalization, and steerability in recommendation-driven environments. Survey evidence associates higher general self-efficacy with fewer emotional symptoms even when negative social-media experiences are considered [1]. Experimental and design work shows that clear explanations and steerable controls are associated with greater perceived transparency and reflective engagement—proximal states for metacognitive regulation [3-5]. Mixed-methods youth studies document perceived benefits of personalization alongside concerns about privacy, safety, trust, and control [2]. Together, these results suggest that platforms can affect both task-specific efficacy and metacognitive processes by altering clarity and user control.

5.2 Similarities and Differences across Methods,

Samples, and Contexts

Across methods, studies converge on the importance of transparency and controllability. Randomized and design studies in educational recommender settings observe higher perceived transparency and strategy reflection when explanations or impact views are present [4]. By contrast, open-platform work emphasizes perceived responsiveness and connectedness rather than direct metacognitive scales [6]. Samples also differ: school-based prototypes typically involve adolescent learners under supervision, while social-media studies rely on broader youth samples. These differences help explain variation in outcomes (trust/strategy reflection vs. connectedness/symptoms) while still pointing to a consistent pathway through clarity and user control.

5.3 Methodological Appraisal

Designs include cross-sectional surveys, randomized experiments, mixed-methods studies, and iterative design programs. Measures span general self-efficacy, negative platform experiences, perceived transparency and trust, perceived algorithm responsiveness, and reflective engagement. Analytic strategies range from regression models to ANOVA and descriptive/qualitative analyses [1,2,4,6]. Limitations recur: short interaction windows, reliance on self-reports, early-stage prototypes, and the scarcity of direct metacognitive instruments in social-media contexts. Some constructs are proximal indicators (e.g., trust, perceived responsiveness) rather than direct measures of planning, monitoring, and evaluation. These constraints qualify causal claims and the generalizability from educational recommenders to open social feeds.

5.4 Practical Implications

Platforms can pair ranking with plain-language explanations that show why items appear and how user inputs shape the feed, and they can provide low-friction controls to steer, reset, mute, and diversify content [3-5]. For schools and caregivers, instruction in algorithmic literacy—how feeds learn, how to diversify inputs, and how to use available controls—can help adolescents sustain planning and monitoring during use [2]. Guidance that supports self-regulation may also reduce compulsive checking and help preserve efficacy beliefs [7,8].

5.5 Research Agenda

Future studies should test transparency and steerability on live social platforms using pre-registered designs, manipulate explanation depth and control granularity, and include validated task-specific efficacy and metacognition scales alongside behavioral indicators. Multi-wave designs are

needed to assess durability of effects. Large-scale analytics can report feature-importance stability, incremental predictive gains, and calibration to substantiate claims about model sensitivity for self-regulatory constructs [10]. Research should also examine moderators such as age, baseline algorithmic literacy, task demands, socio-cultural factors, and context (learning vs. social), and increase the use of mixed-methods to connect mechanism-level observations with lived experience.

5.6 Limitations of This Review

The evidence base is heterogeneous and rapidly evolving. This review synthesizes studies that vary in context, measurement, and design; some rely on proximal indicators rather than direct metacognitive measures, and educational prototypes may not capture open-platform dynamics. The review also focuses on adolescent samples in recent years; work outside this scope was not included. These choices frame the inferences drawn here.

6. Conclusion

Across methods and contexts, the consistent finding is that clarity and user control are central to how recommendation algorithms relate to adolescent outcomes, and this pattern holds across cultural and platform contexts. When ranking is opaque and engagement-optimized, monitoring costs rise and planning and evaluation may decline; when explanations and control options are available, reflective engagement increases, supporting metacognitive regulation and preserving task-specific efficacy. Advancing this field requires field-based experiments, validated instruments for metacognition, and transparent reporting of analytic choices. These steps can link design decisions to measurable benefits for adolescent users without relying on broad restrictions that may undermine agency.

References

- [1] Bonsaksen, T., Schoultz, M., Leung, J., Ruffolo, M., & Thygesen, H. (2023). Negative social media-related experiences and lower general self-efficacy associated with depressive symptoms in Norwegian adolescents: A cross-sectional analysis. *Frontiers in Psychology, 14*, 9853181. <https://doi.org/10.3389/fpsyg.2023.9853181>
- [2] Pérez Vallejos, E. (2021). The impact of algorithmic decision-making processes on young people's online experiences and well-being. *Health Informatics Journal, 27*(4), 1324–1340. <https://doi.org/10.1177/1460458220972750>
- [3] Chaudhry, M. A., Cukurova, M., & Luckin, R. (2022). A transparency index framework for AI in education. *arXiv preprint arXiv:2206.03220*. <https://doi.org/10.48550/arXiv.2206.03220>
- [4] Ooge, J., Dereu, L., & Verbert, K. (2023). Steering recommendations and visualising its impact: Effects on adolescents' trust in e-learning platforms. *arXiv preprint arXiv:2303.00098*. <https://doi.org/10.48550/arXiv.2303.00098>
- [5] Ooge, J., Vanneste, A., Szymanski, M., & Verbert, K. (2025). Designing visual explanations and learner controls to engage adolescents in AI-supported exercise selection. *arXiv preprint arXiv:2412.16034*. <https://doi.org/10.48550/arXiv.2412.16034>
- [6] Taylor, S. H., Bazarova, N. N., & Wohn, D. Y. (2024). The “lonely algorithm” problem: Algorithmic personalization predicts social connectedness on TikTok. *Journal of Computer-Mediated Communication, 29*(5), zmae017. <https://doi.org/10.1093/jcmc/zmae017>
- [7] McAlister, K. L., Rugiero, K. J., & Davidson, T. M. (2024). Social media use in adolescents: Benefits, bans, and emotional regulation recommendations. *JMIR Mental Health, 11*(1), e64626. <https://doi.org/10.2196/64626>
- [8] De, D., Grossman, Z., & Schulman, J. (2025). Social media algorithms and teen addiction: Neurophysiological pathways, reward circuitry, and emotional regulation consequences. *Journal of Adolescent Neuropsychology, 12*(1), 45–62.
- [9] Zhang, Y. (2022). *A review on the development and promotion of metacognition in primary school students. Education Research and Review, 14*, 123–134. [In Chinese]
- [10] Tan, B., Yang, C., & Xiong, Y. (2025). Using machine learning algorithms to predict students' self-efficacy: Evidence from PISA data. *Learning Analytics and Educational Data Science, 7*(2), 123–137.