AcTion: Aceh Nutrition Journal
p-issn 2527-3310; e-issn 2548-5741

July – September, 2026
Literature Review
DOI: http://dx.doi.org/10.30867/action.v11i3.3346
| Pages: 1105 – 1121 |

Gamified and artificial intelligence–enabled mobile health applications for eating behaviour and weight-related outcomes in children and adolescents: a systematic review

Aplikasi kesehatan seluler bergamifikasi dan berbasis kecerdasan buatan terhadap perilaku makan dan luaran terkait berat badan pada anak dan remaja: tinjauan sistematis

Netta Meridianti Putri[1*], Resti Kurnia Triastanti[2], Disyam Adityana Putra[3]

[1] Department of Nutrition, Universitas Tidar, Magelang, Indonesia. E-mail: nettamp@untidar.ac.id
[2] Department of Nutrition, Universitas Tidar, Magelang, Indonesia. E-mail: r.triastanti@untidar.ac.id
[3] PT. Hash Rekayasa Teknologi, Malang, Indonesia. E-mail: disyam@hash.id

***Corresponding Author:**
Department Department of Nutrition, Universitas Tidar, Magelang 56116, Indonesia. E-mail: nettamp@untidar.ac.id

Article History:
Received: July 01, 2026; Revised: July 20, 2026; Accepted: July 30, 2026; Published: September 15, 2026.

Publisher:
Politeknik Kesehatan Aceh
Kementerian Kesehatan RI

© The Author(s). 2026 **Open Access**
This article has been distributed under the terms of the *License Internasional Creative Commons Attribution 4.0*

Abstract

Pediatric overweight and obesity are rising globally, including in Indonesia. Mobile health (mHealth) applications using gamification, artificial intelligence (AI)-based personalization, or both have been proposed for this group, but the evidence is unsynthesised and unclear. This study aimed to synthesize their effects on eating behavior and weight-related outcomes and describe engagement. Methods: The following PRISMA 2020, four databases were searched (January 2016–July 21, 2026) with citation searching. Eligible studies enrolled participants aged 5–19 years, evaluated an mHealth application with gamification and/or AI-based personalization, and reported eating behavior or weight-related outcomes. Screening, extraction, and risk of bias assessment (RoB 2; NIH pre-post tool) were duplicated. The synthesis was narrative. Results: Five studies (four RCTs and one single-arm study; 486 participants) were included. Two interventions used conversational agents only, two used gamification only, and one used both; none was fully integrated. The dietary game increased simulated healthy food choices immediately after play (2.48 vs. 1.10; p<.001; d=1.25). Diet-tracking and gamified lifestyle applications produced no significant dietary changes, and none improved the BMI-SDS or z-score. Engagement metrics were non-comparable and showed a decline. Three RCTs had a high risk of bias. In conclusion, these applications are promising but not yet proven; current evidence is insufficient rather than negative. Adequately powered, registered trials with standardized outcomes and longer follow-ups are needed, including in Indonesia.

Keywords: Adolescent obesity, artificial intelligence, diet adherence, gamification, mHealth

Abstrak

Kelebihan berat badan dan obesitas anak meningkat secara global, termasuk di Indonesia. Aplikasi kesehatan seluler (mHealth) bergamifikasi, berbasis kecerdasan buatan (AI), atau keduanya diusulkan, tetapi buktinya belum dirangkum. Studi bertuuan untuk merangkum pengaruhnya terhadap perilaku makan dan luaran terkait berat badan serta keterlibatan pengguna. Metode, studi menggunakan PRISMA 2020, empat basis data ditelusuri (Januari 2016–21 Juli 2026) disertai penelusuran sitasi. Studi yang memenuhi syarat melibatkan peserta usia 5–19 tahun, mengevaluasi aplikasi mHealth bergamifikasi dan/atau berbasis AI, serta melaporkan luaran perilaku makan atau terkait berat badan. Penapisan, ekstraksi, dan penilaian risiko bias (RoB 2; NIH pre-post) dilakukan ganda. Sintesis bersifat naratif. Hasil, terdapat lima studi (empat RCT, satu lengan tunggal; 486 peserta) disertakan. Dua intervensi hanya memakai agen percakapan, dua hanya gamifikasi, satu keduanya; tidak ada yang terintegrasi penuh. Permainan diet meningkatkan pilihan makanan sehat tersimulasi segera setelah bermain (2,48 vs 1,10; P<0,001; d=1,25); aplikasi pencatatan diet dan gaya hidup bergamifikasi tidak mengubah asupan secara bermakna, dan tidak ada yang memperbaiki BMI-SDS atau z-score. Metrik keterlibatan tidak sebanding dan menurun. Tiga RCT berisiko bias tinggi. Kesimpulan, aplikasi ini menjanjikan namun belum terbukti; bukti saat ini belum memadai, bukan menunjukkan ketiadaan manfaat. Diperlukan uji acak terdaftar berkekuatan memadai dengan luaran terstandar dan tindak lanjut lebih panjang, termasuk di Indonesia.

Kata Kunci: Gamifikasi, kecerdasan buatan, kepatuhan diet, mHealth, obesitas remaja

Introduction

Overweight and obesity in childhood and adolescence are important determinants of non-communicable disease in later life (NCD Risk Factor Collaboration [NCD-RisC], 2024; Zhang et al., 2024). They frequently persist into adulthood (Simmonds et al., 2016) and are associated with cardiometabolic and psychosocial morbidity, and because they arise from interacting dietary, physical-activity, environmental and behavioural factors, early, scalable and sustainable prevention and treatment strategies are a public health priority.

Indonesia is no exception. The 2023 Indonesian Health Survey (Survei Kesehatan Indonesia, SKI 2023) reported that among adolescents aged 13–15 years the prevalence of overweight was 12.1% and of obesity 4.1% (Ministry of Health of the Republic of Indonesia, 2023). The same survey reported that obesity among adults older than 18 years rose from 21.8% in 2018 to 23.4% in 2023; that adult estimate is cited here only as general context for the national trend and is not evidence about adolescents. The adolescent-specific figures provide the appropriate justification for interventions that can reach young people broadly and sustainably.

Mobile health (mHealth) applications are a plausible delivery channel for such interventions, given the high level of mobile device ownership among young people in Indonesia and comparable settings (BPS-Statistics Indonesia, 2023; Kouvari et al., 2022; Melo et al., 2025). Two design strategies are commonly used to sustain uptake and engagement: gamification and artificial intelligence (AI)-based personalisation. These are conceptually distinct from each other and from three related terms with which they are frequently conflated, so this review adopts the following operational definitions. Gamification is the use of game design elements, points, badges, leaderboards, levels, challenges and virtual rewards , within a non-game context such as a health application, where the underlying activity remains a health task (Cheng et al., 2019; Johnson et al., 2016; Sardi et al., 2017). A serious game is a complete game with its own goals, rules and narrative, designed for a purpose beyond entertainment; it is not an application to which game elements have been added, and evidence from the two is not interchangeable. Adaptive learning adjusts content or difficulty to the learner's performance, and such adjustment may be rule-based rather than AI-based. AI-based personalisation is the tailoring of content, timing or feedback by algorithms that learn from user data, including machine-learning-driven recommender systems. Chatbots and conversational agents interact with the user through dialogue and may be either rule-based, using scripted decision trees, or AI-based, using natural language understanding and adaptive response generation; only the latter constitute AI-based personalisation, and primary reports do not always make this distinction explicit.

Evidence of effectiveness is mixed across these approaches. A systematic review of AI chatbots concluded that such interventions may increase physical activity but did not permit definitive conclusions regarding diet or weight management (Oh et al., 2021). In children and adolescents, a meta-analysis found that gamification improved nutritional knowledge but had no significant effect on the body mass index (BMI) z-score (Suleiman-Martos et al., 2021). In adults, weight-loss applications produced modest reductions in weight that attenuated over time, from 2.18 kg at three months to 1.63 kg at 12 months (Chew et al., 2022). A Cochrane review of mHealth smartphone interventions for adolescents and adults with overweight or obesity also reported uncertain effects (Metzendorf et al., 2024). Reviews of gamified health applications have also noted that rewards and feedback are the most commonly implemented elements, but rigorous empirical evaluations of these elements are scarce (Edwards et al., 2016; Sardi et al., 2017).

Taken together, previous studies suggest that digital tools may improve proximal behaviors and knowledge, while their effect on weight-related outcomes remains uncertain, particularly in young people (Kouvari et al., 2022; Melo et al., 2025). What has not been synthesized is the evidence for the specific population of children and adolescents across the full range of these technologies as they actually appear in the literature, gamification alone, AI-based personalization alone, or the two in combination. However, an integrated "AI-driven gamification" intervention is not supported by the primary literature; therefore, this review does not presuppose such integration.

Accordingly, this review addresses a single question, applied consistently throughout: in children and adolescents aged 5–19 years, do mHealth applications that incorporate gamification and/or AI-based personalization, compared with usual care, waitlist, an active comparator, or a non-gamified/non-AI application, improve eating behavior outcomes or weight-related outcomes? User engagement was examined as a secondary descriptive outcome. The degree of integration between AI and gamification components was treated as a characteristic to be described, not as an eligibility requirement.

Methods

Design, Reporting, and Protocol Registration

This systematic review was designed and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines (Page et al., 2021). Given the diversity of study designs, populations, and outcome measures anticipated, a narrative synthesis without meta-analysis was planned a priori.

The review protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO; registration number CRD420261456152, registered in 2026). The registered record corresponds to the review reported here, and no deviations from it arose. The review question, eligibility criteria, and analysis plan were specified before screening began. Readers should consult the PROSPERO record for the registration date when judging the extent to which the outcomes and analyses reported here can be verified as pre-specified.

Eligibility Criteria (PICOS)

Eligibility was defined using the PICOS framework as follows: All criteria were operational and applied without discretionary exceptions.

Population. Children and adolescents aged 5–19 years at enrolment of any weight status. Studies enrolling a mixed-age sample were eligible if the mean age fell within this range or if the results for participants aged 5–19 years were reported separately. Weight status, which was used as an inclusion criterion by the primary study, was recorded together with the diagnostic reference applied (for example, WHO or International Obesity Task Force BMI-for-age cut-offs, or national growth references). Studies restricted to adults, pregnant adolescents, or populations with a primary diagnosis other than overweight/obesity (e.g., type 1 diabetes) were excluded.

Intervention. A mobile health application or mobile game delivered on a smartphone or tablet containing at least one of the following: (i) an explicit gamification element (points, badges, levels, leaderboards, challenges, avatars, or virtual rewards) or (ii) an AI-based personalization component (machine-learning-driven tailoring, adaptive algorithms, or an AI-based conversational agent). Interventions combining application with human support (coach, clinician, or family involvement) were eligible, and the co-intervention was recorded. Web-only platforms without mobile applications, SMS-only interventions without gamification or AI, and wearable devices without application components were excluded.

Comparator. Usual care, waitlist control, no intervention, active non-gamified/non-AI comparator, or alternative application. Single-arm studies without a comparator were eligible only for the descriptive feasibility and engagement objectives and were not used to support conclusions about effectiveness; this is explicitly stated wherever such a study is cited.

Outcomes. At least one of the following criteria was required: primary outcomes: (i) eating behavior outcomes, comprising dietary intake (energy, food group, or nutrient intake measured by recall, record, or food-frequency questionnaire) and dietary adherence to a prescribed pattern; and (ii) weight-related outcomes (BMI, BMI standard deviation score [BMI-SDS], BMI z-score [zBMI], body weight, waist circumference, or body fat percentage). Secondary outcomes included nutrition knowledge, simulated or in-game food choices, and user engagement. These outcome categories were pre-specified and reported separately throughout the study period. In particular, in-game food choices, nutrition knowledge, self-reported fruit and vegetable intake, and measured dietary intake were not treated as equivalent.

Study design. Randomized controlled trials, cluster-randomized trials, non-randomized controlled trials, quasi-experimental studies, and single-arm before-and-after or feasibility studies. Protocols, reviews, editorials, commentaries, modelling studies, and conference abstracts were excluded. Where a review article is cited in this manuscript, it is cited as background literature and is clearly distinguished from the primary studies included.

Other limits. Publications from January 1 2016,, 2016 onwards, to reflect the period in which smartphone-delivered gamified and AI-enabled interventions became widespread. No language restriction was applied at the search stage; the language of the retrieved reports is reported in the Results section.

Information Sources and Search Strategies

PubMed/MEDLINE, Scopus, Web of Science, and ScienceDirect were searched from January 1 2016,, 2016 to July 21, 2026. Forward and backward citation searches of all included reports and relevant reviews were also performed.

The two concept blocks capturing the intervention, gamification, and AI-based personalization, were combined with OR, so that a record addressing either technology was retrieved; the review question did not require both to be present in a single study.

(gamification OR "serious game" OR "game-based" OR "video game" OR "exergam*" OR "artificial intelligence" OR chatbot OR "conversational agent" OR "machine learning" OR "adaptive learning" OR "recommender system") AND ("mobile health" OR mHealth OR "smartphone app*" OR "mobile application" OR "mobile app") AND (adolescen* OR youth OR teen* OR child* OR pediatric OR paediatric) AND (obesity OR overweight OR diet OR eating OR nutrition OR "body weight")

The complete database-specific strategies, including controlled vocabulary (MeSH terms in PubMed/MEDLINE), field tags, limits applied, date each search was run, and number of records retrieved from each source, are provided in Supplementary File 1. The total yield by source is shown in Figure 1.

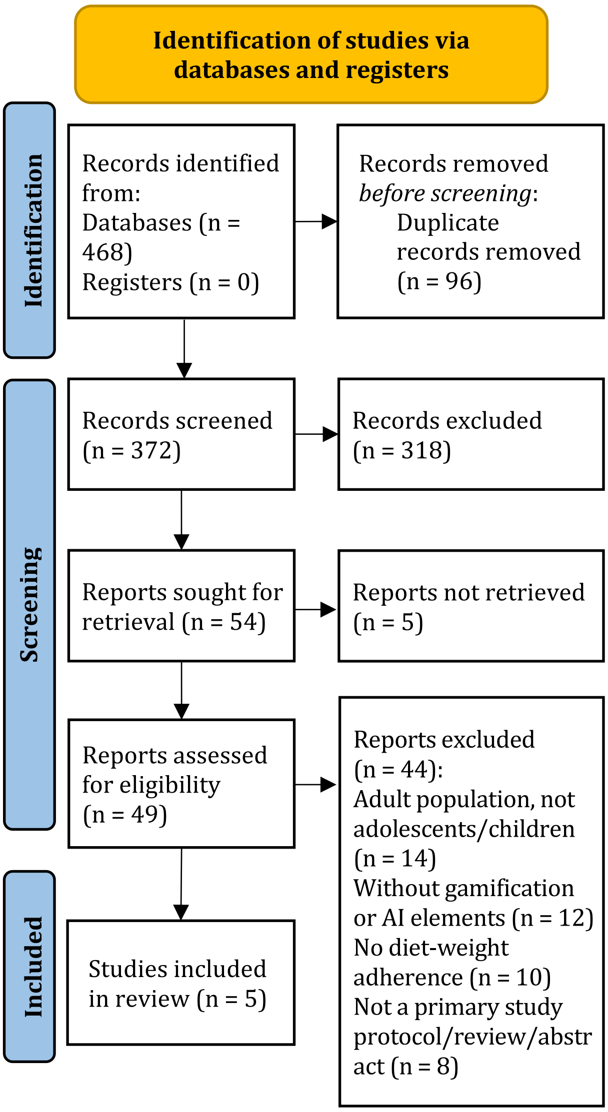

Figure 1. PRISMA 2020 flowchart of the study selection process

Study Selection and Data Extraction

Records were de-duplicated and screened by title and abstract by two reviewers working independently, followed by an independent full-text assessment of potentially eligible reports. Data extraction was performed independently and in duplicate using a piloted, standardized form. Disagreements at every stage were resolved by discussion and consultation with a third reviewer when consensus could not be reached.

The extracted items were author, year, country and setting; study design and trial registration; sample size randomized and analyzed; participant age, sex distribution, weight status, and the diagnostic criteria used; intervention platform, gamification elements, AI components, and their degree of integration; intervention dose and duration; co-interventions and human support; comparator; length of follow-up; attrition; outcome definitions and measurement instruments; and effect estimates with measures of precision where reported. Each intervention was classified as AI-only, gamification-only, or combined AI-plus-gamification, as reported in Table 1.

Risk of Bias Assessment

The risk of bias was assessed independently by two reviewers using design-appropriate instruments. Randomized controlled trials were assessed using the Cochrane risk-of-bias tool for randomized trials, version 2 (RoB 2), across its five domains: (1) bias arising from the randomization process; (2) bias due to deviations from intended interventions; (3) bias due to missing outcome data; (4) bias in the measurement of the outcome; and (5) bias in the selection of the reported result. Each domain was rated as low risk, some concerns, or high risk, and an overall judgement was derived according to the RoB 2 algorithm. The single-arm before-after study was assessed using the National Institutes of Health (NIH) Quality Assessment Tool for Before-After (Pre-Post) Studies With No Control Group and rated as good, fair, or poor. Disagreements were resolved through discussion.

Risk-of-bias judgements were incorporated into the synthesis in three explicit ways. First, results from studies at high risk of bias and from the single-arm study are not used to support statements about effectiveness; they are reported descriptively and labelled as such. Second, the direction and likely magnitude of bias in each domain are discussed alongside the corresponding results. Third, the overall certainty of the body of evidence for each outcome was appraised qualitatively using the GRADE domains (risk of bias, inconsistency, indirectness, imprecision, and publication bias) and is reported at the end of the Results. The full risk of bias results are presented in Tables 2 and 3.

Data Synthesis

A meta-analysis was not performed. The included studies differed substantially in population (children with no weight-status criterion to adolescents with obesity), intervention (single-session game to a 5.5-month blended intervention), comparator (waitlist, active application, board game), and outcome definition (in-game choice, self-reported intake score, BMI-SDS). Pooling across these differences would have produced a summary estimate without coherent interpretation. Results are therefore grouped by pre-specified outcome category and, within each category, by intervention type, and effect estimates are reported as given in the primary reports.

Result and Discussion

Study Selection

The search identified 468 records (457 from databases and 11 from other sources). After 96 duplicates were removed, 372 records were screened based on their titles and abstracts, of which 318 were excluded. A total of 54 reports were sought for full text, 5 were not retrieved, and 49 were assessed for their eligibility. After 44 reports were excluded for various reasons, five studies were included in this review. The selection flowchart is shown in Figure 1.

Two of the excluded reports lie close to the eligibility boundary and are named here for transparency purposes. A randomized trial of the CommitFit gamified application in 30 adolescent–caregiver dyads was excluded under the outcome criterion because the published report presented only acceptability, engagement, and motivation results and stated that eating behavior and weight-related results are reserved for a separate report (Henry et al., 2025); the parent trial is best regarded as a study awaiting classification, and its outcome paper should be sought in any future update of this review. A multifactorial trial of an SMS-delivered chatbot education program in 7,890 Danish adolescents was excluded under the intervention criterion because the program was SMS-only rather than application-based, its dialogue was scripted rather than AI-driven, and tailoring was achieved through the factorial design rather than algorithmically (Bjerregaard et al., 2024). Both exclusions follow directly from the pre-specified criteria; the second is relevant to the interpretation of the findings and is discussed below.

Table 1. Summary of studies included in the review

Author (year), country	Design, registration, n randomised / analysed	Population: age, sex, weight status and criterion	Intervention: classification, features, dose, duration	Comparator; follow-up; attrition	Outcomes and effect estimates
Stephens et al. (2019), United States	Single-arm before-after feasibility study; no comparators. This study was exempt from ethical approval. n = 23 enrolled / 23 analysed.	Adolescents enrolled in a pediatric obesity and prediabetes program. Age and sex distribution were not reported in a form permitting extraction. Weight status: program entry (obesity/prediabetes); explicit BMI criterion not reported.	Classification: AI only. Tess, an AI behavioral coaching chatbot delivered by SMS and Facebook Messenger; adaptive conversation, psychoeducation, and goal support. No gamification elements were described. Duration: 10–12 weeks.	None (single-arm). Follow-up: end of programme. Attrition not reported.	Eating behavior: Not measured. Weight-related: not measured as a study outcome. Engagement (descriptive): 4,123 messages exchanged; positive progress towards goals reported in 81% of interactions; chatbot rated helpful in 96% of ratings. No between-group inferences were possible.
Stasinaki et al. (2021), Switzerland	Two-arm parallel RCT (1:1), not blinded. Registration: Not stated in the primary report. n = 41 randomized / 31 analyzed (18 intervention, 13 control).	Age 10–18 years. Sex distribution: see the primary report. Weight status: BMI > 97th percentile or BMI > 90th percentile with risk factors or comorbidities (national percentile reference).	Classification: AI + gamification. PathMate2: daily conversational agent counselling via app, with game characters, challenges, virtual rewards, avatars, and a progress dashboard, combined with four on-site standardized counselling visits. The AI did notthe game mechanics. Duration 5.5 months.	Multi-component behavior-changing program (active comparator). Follow-up 12 months. Attrition ~24% (41 → 31 analyzed).	Eating behavior: dietary adherence was not reported as a quantified outcome. Weight-related: BMI-SDS did not decrease significantly in the intervention group at either time point; BMI-SDS decreased significantly in the control group at 5.5 months but not at 12 months. Median baseline BMI-SDS 2.61. Body fat decreased significantly in the intervention group, and muscle mass, strength, and agility improved in both groups at 12 months. Engagement (descriptive): mean daily app usage rate was 71.5%.
Kato-Lin et al. (2020), India	Two-arm RCT, individually randomized within schools; not blinded. Registration: not reported. n = 104 randomized; data from one school excluded from some	Age 10–11 years (children/pre-adolescents). Sex distribution was reported in the primary report. Weight status: no weight status entry criterion; general school population.	Classification: gamification only. fooya!, a dietary mobile game with implicit and adaptive learning (levels, in-game food choices, and nutrition facts). No AI-based personalization was described. Dose: Two 20-minute	Uno, a board game without dietary content (active comparator). Follow-up: Immediately after each session. Attrition: not applicable (single-visit	Eating behavior (simulated, not dietary intake): number of healthy items chosen from a set of six offered immediately after play — 2.48 (intervention) vs. 1.10 (control); P < .001; Cohen's d = 1.25. Healthy foods correctly identified 7.3 vs 6.94; P = .048; Cohen d = 0.25.

Author (year), country	Design, registration, n randomised / analysed	Population: age, sex, weight status and criterion	Intervention: classification, features, dose, duration	Comparator; follow-up; attrition	Outcomes and effect estimates
	analyses.		play sessions.	design); data from one school excluded.	Weight-related: Not measured. Engagement (descriptive): mean 15 levels played across two sessions (range, 2–23).
Tugault-Lafleur et al. (2023), Canada	Two-arm parallel RCT; not blinded. Registration: ClinicalTrials.gov; published protocol is available. n = 214 adolescent–parent pairs were randomized.	Age 10–17 years. Sex distribution was reported in the primary report. Weight status: overweight or obese (BMI-for-age criterion as defined in the primary report).	Classification: gamification only. Aim2Be, a gamified lifestyle application (points, challenges, interactive content) for adolescents and parents, with live coach support in the intervention group. No AI-based personalization was described. Duration: 3 months (primary comparison); 6 months of access.	Waitlist control (access to the app after three months). Follow-up at 3 and 6 months. Attrition: see the primary report.	Eating behavior: No significant between-group differences in self-reported lifestyle behaviors, including fruit and vegetable intake and sugary drink intake, at 3 months. Weight-related: No significant between-group difference in zBMI at 3 months; app use did not modify outcomes. Engagement (descriptive): declined over time; the coach-supported group was more active.
Shatwan et al. (2023), Saudi Arabia	Two-arm RCT; not blinded. Registration: ClinicalTrials.gov, NCT05692765 (registered near the end of the study period). n = 104 randomized (55 intervention, 49 control subjects).	Age 13–18 years. Sex distribution was reported in the primary report. Weight status: no weight status entry criterion; general adolescent population in Jeddah.	Classification: Neither gamification nor AI is a principal mechanism. MyPlate is a smartphone application for tracking and goal-setting around fruit and vegetable intake (participants selected three of seven fruit goals and three of seven vegetable goals). No gamification elements or AI components were described by the authors. Duration 6 weeks.	No-app control (questionnaires only). Follow-up 6 weeks. Attrition: see the primary report.	Eating behavior (self-reported consumption scores; within-group comparisons): intervention group fruit 1.48 (SD 0.99) to 1.70 (SD 1.11), P = .31; vegetables 1.50 (SD 0.97) to 1.43 (SD 1.03), P = .30. The control group's fruit intake increased from1.15 (SD 0.68) to 1.64 (SD 0.98), P = .01; and vegetable intake increased from 1.44 (SD 0.97) to 1.55 (SD 0.90), P = .54. Between-group estimates with confidence intervals were not reported in this study. Weight-related: not measured. Engagement: Not reported.

AI = artificial intelligence; RCT = randomized controlled trial; BMI = body mass index; BMI-SDS/zBMI = BMI standard deviation score for age and sex

Table 2. Risk of bias in the four included randomised controlled trials (Cochrane RoB 2)

Study	D1 Randomisation process	D2 Deviations from intended interventi-ons	D3 Missing outcome data	D4 Measurement of the outcome	D5 Selection of the reported result	Overall	Principal reason for the judgement
Stasinaki et al. (2021)	Some concerns	Some concerns	High	Some concerns	Some concerns	High	41 randomised but 31 were analyzed (~24% loss) in an unblinded trial; allocation concealment was not fully described; and outcome assessors were not blinded.

Study	D1 Randomisation process	D2 Deviations from intended interventi-ons	D3 Missing outcome data	D4 Measurement of the outcome	D5 Selection of the reported result	Overall	Principal reason for the judgement
Kato-Lin et al. (2020)	Some concerns	Some concerns	Some concerns	High	Some concerns	High	The outcome was an unblinded simulated choice task administered immediately after play; the intervention game was developed by one of the authors; and data from one school were excluded from some analyses.
Tugault-Lafleur et al. (2023)	Low	Some concerns	Some concerns	Some concerns	Low	Some concerns	The study was registered with a published protocol and adequate randomization; however, participants were unblinded, and the behavioral outcomes were self-reported, and anthropometry was measured at home.
Shatwan et al. (2023)	Some concerns	Some concerns	Some concerns	High	Some concerns	High	Unblinded, self-reported consumption scores; trial registered near the end of the study period; results presented as within-group rather than between-group comparisons.

D1–D5 = RoB 2 domains. Judgements were made independently by two reviewers and reconciled through discussion. Where a primary report did not describe a procedure (for example, allocation concealment), the domain was rated “some concerns” rather than “low risk,” in line with RoB 2 guidance.

Table 3. Quality assessment of the single-arm before-after study (NIH Quality Assessment Tool for Before-After [Pre-Post] Studies With No Control Group)

Study	Assessment against the NIH criteria	Overall rating
Stephens et al. (2019)	The objective was clearly stated, and the eligibility criteria were pre-specified. Participants representative of the clinic population of interest: partially recruited from a single pediatric obesity and prediabetes programme. All eligible participants were enrolled: not reported. Sample size sufficient to detect a change: no; n = 23, with no sample size justification. Intervention was clearly described and consistently delivered: yes. Outcome measures pre-specified, valid, and reliable: no; outcomes were interaction counts and user ratings, not validated dietary or anthropometric measures. Outcome assessors blinded: No. Loss to follow-up was 20% or less, and accounted for: not reported. Statistical methods examined pre-post change with significance testing: not applicable; descriptive only. Multiple outcome measurements before and after the intervention: No.	Poor (for the purpose of inferring effectiveness; adequate for its stated feasibility objective)

Characteristics of Included Studies

The five included studies comprised four RCTs (Kato-Lin et al., 2020; Shatwan et al., 2023; Stasinaki et al., 2021; Tugault-Lafleur et al., 2023) and one single-arm feasibility study (Stephens et al., 2019), published between 2019 and 2023, and conducted in India (Kato-Lin et al., 2020), Switzerland (Stasinaki et al., 2021), Canada (Tugault-Lafleur et al., 2023), Saudi Arabia (Shatwan et al., 2023), and the United States (Stephens et al., 2019). Kato-Lin et al. (2020) conducted their study in schools in India, with data analysis led from the United States. No studies have been conducted in Indonesia or any other Southeast Asian countries. The sample sizes ranged from 23 to 214, with a total of 486 participants.

The populations differed significantly in terms of weight status. Two studies enrolled participants with overweight or obesity as the entry criterion (Stasinaki et al., 2021; Tugault-Lafleur et al., 2023). One enrolled adolescent was already attending a pediatric obesity and prediabetes program (Stephens et al., 2019). Two enrolled participants had no weight-status criterion: schoolchildren aged 10–11 years (Kato-Lin et al., 2020) and adolescents aged 13–18 years from the general school population

(Shatwan et al., 2023). Therefore, the body of evidence does not uniformly speak to adolescents with obesity, and the review title and conclusions have been aligned with this.

The interventions were classified as follows: AI-only: Tess behavioral coaching chatbot (Stephens et al., 2019). Combined AI plus gamification: PathMate2, a daily conversational agent delivered through an application with game characters, challenges, virtual rewards, and a progress dashboard, plus face-to-face counselling (Stasinaki et al., 2021). Gamification-only: fooya!, a dietary mobile game with implicit and adaptive learning components (Kato-Lin et al., 2020), and Aim2Be, a gamified lifestyle application with points and challenges, plus live coach support (Tugault-Lafleur et al., 2023). Neither gamification nor AI as a principal mechanism: MyPlate, a diet-tracking application with goal-setting but no gamification elements or AI component described by the authors (Shatwan et al., 2023); this study was retained as a non-gamified, non-AI active comparator case, and its results were interpreted accordingly. No study has evaluated a fully integrated system in which AI adapts the game mechanics. The full characteristics are presented in Table 1.

Eating-behaviour outcomes

Eating behavior outcomes were measured using instruments that are not interchangeable and are therefore reported separately by measurement type.

Simulated (in-game or immediate post-play) food choices. In a single study, children aged 10–11 years who played a dietary mobile game with implicit learning chose more healthy items from a set offered immediately after play than children who played a board game (2.48 vs 1.10; $P < .001$; Cohen's $d = 1.25$) and identified healthy foods slightly more accurately (7.3 vs 6.94; $P = .048$; Cohen's $d = 0.25$) (Kato-Lin et al., 2020). This outcome is a laboratory-style choice task administered a few minutes after the intervention. It is a measure of immediate, situation-specific preference rather than habitual dietary intake, and the large effect size should not be interpreted as evidence of a change in diet.

Self-reported intake of specific food groups. A diet-tracking application produced no significant within-group changes in fruit ($P = 0.31$) or vegetable ($P = 0.30$) consumption scores over six weeks; the control group showed a significant increase in fruit consumption score ($P = 0.01$), which the authors did not attribute to the intervention, and no between-group estimates with confidence intervals were reported (Shatwan et al., 2023). A gamified lifestyle application with coach support produced no significant between-group differences in self-reported lifestyle behaviors, including fruit and vegetable and sugary drink intakes, at three months (Tugault-Lafleur et al., 2023).

Measured dietary intake and adherence. No included study measured dietary intake using 24-hour recall, food record, or a validated food frequency questionnaire, and no included study measured adherence to a prescribed dietary pattern. No included study measured dietary intake using 24-hour recall, food record, or a validated food frequency questionnaire, and no included study measured adherence to a prescribed dietary pattern. Therefore, dietary intake as actually consumed is essentially unmeasured in the available evidence.

Taken together, the eating behavior evidence comprises one positive result on an immediate simulated-choice task in children without a weight-status criterion, alongside two studies that found no measurable change in self-reported intake. These non-significant findings should not be read as demonstrating the absence of effect: both studies were small and short (six weeks and three months), and neither reported a between-group estimate with a confidence interval able to exclude a clinically meaningful difference.

Weight-Related Outcomes

Only two studies measured weight-related outcomes. In adolescents with obesity, a combined conversational agent and gamified intervention did not produce a significant decrease in BMI-SDS at 5.5 or 12 months, although body fat decreased significantly in the intervention group and muscle mass, strength, and agility improved in both groups (Stasinaki et al., 2021). A gamified lifestyle application produced no significant between-group difference in zBMI at three months compared with a waitlist control (Tugault-Lafleur et al., 2023). No included study reported a statistically significant improvement in anthropometric indices attributable to the application.

With only two contributing studies, one of which lost approximately a quarter of its randomized participants to analysis, and neither powered for this comparison, the evidence base for weight-related outcomes is too sparse and imprecise to support a conclusion in either direction. How these findings sit alongside the wider literature is considered in the Discussion.

Engagement

Engagement was the most favorably reported outcome, although it was measured with metrics that cannot be compared across studies and was not consistently high across them. A single-arm feasibility study of an AI chatbot recorded 4,123 message exchanges, with positive progress towards goals reported in 81% of interactions and the chatbot rated helpful in 96% of ratings (Stephens et al., 2019); this study had no comparator, and its engagement figures could not be benchmarked. A conversational agent application achieved a mean daily usage rate of 71.5% over 5.5 months (Stasinaki et al., 2021), a different construct measured on a different scale. In the largest trial, engagement declined over the intervention period and was higher in the coach-supported group (Tugault-Lafleur et al., 2023). One study reported only descriptive game play patterns (Kato-Lin et al., 2020), and one did not report engagement at all (Shatwan et al., 2023).

Risk of Bias in Included Studies

Risk-of-bias assessments are reported in full in Tables 2 (RoB 2, four RCTs) and 3 (NIH pre-post tool, one single-arm study). One trial raised some concerns (Tugault-Lafleur et al., 2023), and three trials were judged to be at high risk of bias (Kato-Lin et al., 2020; Shatwan et al., 2023; Stasinaki et al., 2021). The single-arm feasibility study was rated poor quality for the purpose of inferring effectiveness, which is consistent with its design and stated feasibility objective (Stephens et al., 2019).

Three features drove these judgements. First, none of the trials could blind participants or personnel, which is important for the self-reported behavioral outcomes that constitute most of the eating behavior evidence. Second, missing outcome data were substantial in one trial, in which 41 participants were randomized but 31 contributed to the analysis; loss of this magnitude in an unblinded trial can plausibly change the direction of a null result. Third, outcome measurement was vulnerable to bias in the studies whose results were most favorable: the simulated food-choice task was administered immediately after an unblinded intervention, and the game evaluated was developed by one of the study authors, a conflict of interest declared in the primary report.

Because the one positive eating behavior result comes from a study at high risk of bias, and both weight-related results come from studies at high risk or with some concerns, no effectiveness conclusion in this review rests on any single study. Therefore, the synthesis was deliberately cautious.

Certainty of the evidence

Using the GRADE domains qualitatively, the certainty was very low for all outcomes. For eating behavior outcomes, certainty was downgraded for risk of bias (the only positive result came from a study at high risk of bias), for inconsistency (one large positive effect alongside two null results), for indirectness (the positive result was an immediate simulated-choice task in children without a weight-status criterion, not dietary intake in adolescents with obesity), and for imprecision (small samples; between-group estimates with confidence intervals largely unreported). For weight-related outcomes, certainty was downgraded for risk of bias, imprecision (two studies, both null, neither powered for this comparison), and indirectness. Publication bias could not be formally assessed in five studies, but it could not be excluded. These ratings, rather than the individual P-values, are the appropriate basis for interpreting this review.

Summary of the Evidence

This review synthesized five primary studies that evaluated mHealth applications with gamification and/or AI-based personalization in children and adolescents. Three points emerge, each qualified by the certainty of the underlying evidence.

First, engagement indicators were favorable in the two studies using conversational agents, but these were measured with non-comparable metrics, one study had no comparator, and engagement declined over time in the largest trial. These are preliminary observations from a small, heterogeneous set of studies. Second, one gamified intervention improved an immediate simulated food-choice outcome in children, while self-reported intake did not change in the two studies that measured it. Third, no intervention improved anthropometric outcomes, but only two studies measured one, and both were underpowered for that comparison. The overall picture is not that these interventions are ineffective; rather, the evidence is insufficient to determine whether they work.

Consistency with the Wider Literature

The null weight-related findings of this review are consistent with evidence from outside the eligibility boundary (Kouvari et al., 2022; Li et al., 2025). A meta-analysis in children and adolescents found no significant effect of gamification on BMI z-score (Suleiman-Martos et al., 2021), and a Cochrane review of mHealth smartphone interventions for adolescents and adults with overweight or obesity reported uncertain effects (Metzendorf et al., 2024).

More directly, a multifactorial randomized trial of an SMS-delivered chatbot education programme in 7,890 Danish adolescents, excluded from this review because the intervention was SMS-only rather than app-based, its dialogue was scripted rather than AI-driven, and tailoring was achieved through the factorial design rather than algorithmically, found no demonstrable effect on either BMI z-score or an abbreviated Healthy Eating Index at 6 or 18 months (Bjerregaard et al., 2024).

The null result obtained in a sample more than 15 times larger than the combined sample of the five included studies suggests that the null findings reported here are unlikely to be explained by small sample size alone, although the intervention differed in delivery mode and intensity. All of these are cited as background literature, and none is an included study.

Why Engagement and Weight-Related Outcomes May Diverge

If the engagement signal is real, several factors could explain why it has not yet been reflected in anthropometric changes. Anthropometric changes in young people with obesity require a sustained energy deficit, and most included interventions lasted from a single session to approximately six months, which limits the opportunity to detect such changes (Shatwan et al., 2023; Stasinaki et al., 2021; Tugault-Lafleur et al., 2023).

Engagement is a proximal process measure rather than a clinical outcome, and a high interaction volume does not indicate a change in the energy intake (Milne-Ives et al., 2024; Perski et al., 2017). Most applications operated as standalone or lightly supplemented interventions, whereas effective adolescent obesity management is typically multicomponent, involving diet, physical activity, behavioral support, and family engagement. Finally, engagement declines over time; therefore, the intervention dose received is often lower than intended (Tugault-Lafleur et al., 2023). These remain hypotheses, and none were tested in the included studies.

AI components in the Included Studies

Where AI was present, it took the form of a conversational agent providing adaptive support, psychoeducation, and goal feedback (Stasinaki et al., 2021; Stephens et al., 2019). Only these two studies described AI components. Kato-Lin et al. (2020) evaluated a game with implicit and adaptive learning components, which the authors did not describe as AI-based, and Shatwan et al. (2023) evaluated a tracking application with no conversational or adaptive component; neither is classified as a conversational-AI intervention.

The theoretical advantages of conversational agents, such as continuous availability, scalability, and contact outside clinic hours for young people who face barriers to in-person care, are frequently cited; however, they were not tested against these outcomes in the included studies. A systematic review of AI chatbots concluded that such interventions may improve physical activity, while the evidence for diet and weight management remained inconclusive, and noted that most of the included studies involved adults in high-income countries (Oh et al., 2021). This review is background literature and is not one of the five primary studies included here. When conversational AI is deployed with adolescents, deliberate safety design for unintended or inappropriate responses and explicit framing of the chatbot as an adjunct to rather than a replacement for clinical care have been recommended (Thompson & Baranowski, 2019).

Gamification Elements and Their Attribution

The gamification elements can be linked to specific interventions in this review as follows. PathMate2 uses game characters, challenges, virtual rewards, avatars, and a progress dashboard alongside its conversational agent (Stasinaki et al., 2021). Aim2Be uses points, challenges, and interactive content (Tugault-Lafleur et al., 2023). fooya! was not a gamified application but a complete serious game, with levels, in-game food choices, and embedded nutrition facts, and should be described as such rather than grouped with applications to which game elements were added (Kato-Lin et al., 2020). MyPlate uses goal-setting and self-monitoring without game elements (Shatwan et al., 2023). Mechanisms beyond those described in the primary reports, such as leaderboards or social competition, were not evaluated in any of the included studies and are not attributed here.

This distribution is consistent with that of the wider literature. A content analysis of commercially available health applications found that only 4% (64 of 1,680) incorporated gamification, and that the most frequently implemented behavior change techniques were feedback and monitoring, reward, and goals and planning (Edwards et al., 2016). A review of gamification in e-health similarly found rewards and feedback to be the most common elements, while noting that they had rarely undergone rigorous empirical evaluation (Sardi et al., 2017). A recurring critique is that gamification relies heavily on extrinsic motivators, whose effects attenuate over time (Cheng et al., 2019; Johnson et al., 2016). Designs intended to produce lasting change are argued to need to support intrinsic motivation, competence, autonomy, and relatedness, rather than points and badges alone; this argument is theoretically motivated and was not tested in any of the included studies.

Integration of AI and Gamification

Although "AI-driven gamification" is often discussed as a distinct class of intervention, it was not represented in the retrieved evidence. In one study combining both, the conversational agent and game elements operated in parallel: the AI did not adapt challenges, rewards, or difficulty, and the authors did not test or discuss any synergy between the two components (Stasinaki et al., 2021).

Adapting game challenges and feedback to a user's stage of change, food preferences, and nutritional requirements is a coherent design proposal, and reviews of recommender systems in weight-management applications indicate that tailoring features are increasingly common, while behavioral theory remains under-applied and engagement remains low (de Castro et al., 2025). Nevertheless, this remains an untested proposal, and this review cannot support or refute it as a hypothesis. Future trials that deliberately manipulate the degree of integration are needed to answer this question.

Measuring Eating Behaviour: A Barrier to Synthesis

The most consequential methodological finding of this review is that the included studies measured different constructs under different headings. In-game food choice measured minutes after play (Kato-Lin et al., 2020), a self-reported fruit and vegetable consumption score (Shatwan et al., 2023), and a self-reported composite of lifestyle behaviors (Tugault-Lafleur et al., 2023) are not interchangeable, and none of these measures habitual dietary intake or adherence to a dietary prescription.

Knowledge gains and correct choices within a game environment can be produced within a single session and measured immediately, which tends to yield larger short-term effects than measures of actual intake (Kato-Lin et al., 2020; Suleiman-Martos et al., 2021). Self-reported measures are also vulnerable to recall and social desirability bias (Ravelli & Schoeller, 2020). Agreement on a core outcome set, anchored on measured intake, validated diet quality scores, or objective biomarkers, with a pre-specified taxonomy separating intake, adherence, knowledge, simulated choice, and self-reported behavior, would make future studies comparable and would allow quantitative synthesis.

Moderating Factors

Two intervention characteristics appear to co-vary with outcomes, although with five studies, this can only be described, not tested. Longer and more intensive interventions were those in which any physiological change was observed, whereas the shortest study was neutral (Shatwan et al., 2023).

Human support also appeared relevant: in the gamified lifestyle trial, the coach-supported arm was more active, although the between-group zBMI difference remained non-significant (Tugault-Lafleur et al., 2023), and the conversational agent intervention delivered alongside face-to-face counselling was accompanied by improvements in body composition and physical capacity (Stasinaki et al., 2021). These observations are consistent with the use of mobile applications for positioning as one component of a multicomponent package rather than as a substitute for professional support or family involvement. However, they do not provide evidence of a causal moderating effect (Kouvari et al., 2022; Li et al., 2025).

Comparison with Adult Evidence

Evidence from adults provides context. In adults with overweight or obesity, weight-loss applications produced a pooled reduction of 2.18 kg at three months, attenuating to 1.63 kg at 12 months (Chew et al., 2022).

This pattern of a small effect that attenuates over time suggests that application-delivered interventions are best positioned as a sustainable component of care rather than as a stand-alone solution, even when they are effective. Adolescents differ in the dynamics of motivation, autonomy, and environmental influence from family, school, and peers; therefore, adult findings should not be assumed to transfer without adaptation (Hargreaves et al., 2022).

Theoretical grounding and personalisation

A recurring opportunity in this literature lies in strengthening its grounding in behavior change theory and involving subject-matter experts in application development. A scoping review of recommender systems in weight-management mHealth interventions found that personalization features and tailored messaging are increasingly common, while the application of behavioral theory remains limited, engagement is measured in fewer than half of studies, and there are inconsistencies between the authors' domain expertise and the intervention content (de Castro et al., 2025).

AI-based personalization embedded in a defensible theoretical framework, integrating self-monitoring, goal setting, and tailored feedback, is a plausible route to narrowing the gap between engagement and clinical outcomes, but it is a design hypothesis rather than a demonstrated mechanism.

Implications for the Indonesian Context

No included study was conducted in Indonesia or in a comparable middle-income Southeast Asian setting; therefore, the following are implications for design and future research rather than the findings of this review.

Two population-specific Indonesian figures frame the study. First, overnutrition is already established in Indonesian adolescents: SKI 2023 reported an overweight prevalence of 12.1% and an obesity prevalence of 4.1% among those aged 13–15 years (Ministry of Health of Indonesia, 2023). Second, the delivery channel is available: Statistics Indonesia reported that 96.28% of young people aged 16–30 years who used the internet in the preceding three months did so using a mobile telephone, and that 94.16% of this age group had accessed the internet (BPS-Statistics Indonesia, 2023). Nationally, 72.78% of the population aged 5 years and older accessed the Internet in 2024, and 68.65% owned a mobile telephone (BPS-Statistics Indonesia, 2025).

On that basis, applications developed for Indonesian adolescents would plausibly need to: (a) be grounded in behavior change theory, with nutritionists and other health professionals involved in content design; (b) reflect local food and cultural context, aligning recommendations with the national balanced nutrition guidance ("*Isi Piringku*"), using local food examples and an Indonesian-language interface (Agustina et al., 2022; Prafiantini et al., 2025; Putri et al., 2025; Rachmi et al., 2021; Rahayu et al., 2021; Roshita et al., 2021); (c) involve families and, where feasible, schools, consistent with the observation that human support accompanied the more favorable results in this review; and (d) be positioned as a complement to, not a substitute for, nutrition services, with success framed in terms of eating behavior, nutrition literacy, and sustainable habits rather than weight loss alone. Each of these is a design recommendation requiring evaluation and is not a conclusion supported by the included studies.

Conclusion

Gamified and AI-enabled mHealth applications are a promising but largely untested approach to supporting eating behavior and healthy weight in children and adolescents. The five small, heterogeneous studies available offer early encouraging signals: a serious game improved immediate food choices in 10–11 year olds, and engagement with conversational agent applications was high in the two studies that assessed it. However, the evidence cannot yet show whether these early signals translate into sustained changes in dietary intake or body weight because self-reported intake did not change, the two studies measuring anthropometry found no improvement, and engagement was assessed in ways that cannot be compared across studies.

This is a gap in the evidence, not a verdict on the interventions, and it applies to short-duration programmes in 10–18 year olds in high- and middle-income settings. Two questions remain: whether AI and gamification are more effective when genuinely integrated rather than delivered in parallel, as in every study to date, and how these tools perform in Indonesia, where searching in Indonesian identified no eligible study. Realizing the potential of this approach will require adequately powered registered trials with a core outcome set anchored on measured intake and anthropometry, follow-up beyond 12 months, and explicit reporting of intervention components.

Acknowledgements

The authors thank LPPM Universitas Tidar for their institutional support

References

Agustina, R., Febriyanti, E., Putri, M., Martineta, M., Hardiany, N. S., Mustikawati, D. E., Hanifa, H., & Shankar, A. H. (2022). Development and preliminary validity of an Indonesian mobile application for a balanced and sustainable diet for obesity management. *BMC Public Health, 22*(1), Article 1221. https://doi.org/10.1186/s12889-022-13579-x

Bjerregaard, A. A., Zoughbie, D. E., Hansen, J. V., Granström, C., Strøm, M., Halldórsson, Þ. I., Meder, I. K., Willett, W. C., Ding, E. L., & Olsen, S. F. (2024). An SMS chatbot digital educational program to increase healthy eating behaviors in adolescence: A multifactorial randomized controlled trial among 7,890 participants in the Danish National Birth Cohort. *PLOS Medicine, 21*(6), Article e1004383. https://doi.org/10.1371/journal.pmed.1004383

BPS–Statistics Indonesia. (2023). *Statistik pemuda Indonesia 2023 [Youth statistics of Indonesia 2023].* https://www.bps.go.id/id/publication/2023/12/29/18781f394974f2cae5241318/statistik-pemuda-indonesia-2023.html

BPS–Statistics Indonesia. (2025). *Statistik telekomunikasi Indonesia 2024 [Telecommunication statistics in Indonesia 2024].* https://www.bps.go.id/id/publication/2025/08/29/beaa2be400eda6ce6c636ef8/statistik-telekomunikasi-indonesia-2024.html

Cheng, V. W. S., Davenport, T., Johnson, D., Vella, K., & Hickie, I. B. (2019). Gamification in apps and technologies for improving mental health and well-being: A systematic review. *JMIR Mental Health, 6*(6), Article e13717. https://doi.org/10.2196/13717

Chew, H. S. J., Koh, W. L., Ng, J. S. H. Y., & Tan, K. K. (2022). Sustainability of weight loss through smartphone apps: Systematic review and meta-analysis on anthropometric, metabolic, and dietary outcomes. *Journal of Medical Internet Research, 24*(9), Article e40141. https://doi.org/10.2196/40141

de Castro, B. A., Levens, S. M., Sullivan, M., & Shaw, G., Jr. (2025). Recommender systems use in weight management mHealth interventions: A scoping review. *Obesity Reviews, 26*(3), Article e13863. https://doi.org/10.1111/obr.13863

Edwards, E. A., Lumsden, J., Rivas, C., Steed, L., Edwards, L. A., Thiyagarajan, A., Sohanpal, R., Caton, H., Griffiths, C. J., Munafò, M. R., Taylor, S., & Walton, R. T. (2016). Gamification for health promotion: Systematic review of behaviour change techniques in smartphone apps. *BMJ Open, 6*(10), Article e012447. https://doi.org/10.1136/bmjopen-2016-012447

Hargreaves, D., Mates, E., Menon, P., Alderman, H., Devakumar, D., Fawzi, W., Greenfield, G., Hammoudeh, W., He, S., Lahiri, A., Liu, Z., Nguyen, P. H., Sethi, V., Wang, H., Neufeld, L. M., & Patton, G. C. (2022). Strategies and interventions for healthy adolescent growth, nutrition, and development. *The Lancet, 399*(10320), 198–210. https://doi.org/10.1016/S0140-

[6736(21)01593-2](https://doi.org/10.1016/S0140-6736(21)01593-2)

Henry, N., Montgomery, E., Ghosh, P., Bosworth, K. T., Lim, C., Ghosh, J., Popescu, M., Kimchi, K., Guo, C., Smith, J., & Braddock, A. (2025). Caregiver inclusion influence on adolescent acceptance and engagement of an mHealth app: A randomized controlled trial. *mHealth, 11*, Article 57. https://doi.org/10.21037/mhealth-24-97

Johnson, D., Deterding, S., Kuhn, K.-A., Staneva, A., Stoyanov, S., & Hides, L. (2016). Gamification for health and wellbeing: A systematic review of the literature. *Internet Interventions, 6*, 89–106. https://doi.org/10.1016/j.invent.2016.10.002

Kato-Lin, Y.-C., Kumar, U. B., Sri Prakash, B., Prakash, B., Varadan, V., Agnihotri, S., Subramanyam, N., Krishnatray, P., & Padman, R. (2020). Impact of pediatric mobile game play on healthy eating behavior: Randomized controlled trial. *JMIR mHealth and uHealth, 8*(11), Article e15717. https://doi.org/10.2196/15717

Kouvari, M., Karipidou, M., Tsiampalis, T., Mamalaki, E., Poulimeneas, D., Bathrellou, E., Panagiotakos, D., & Yannakoulia, M. (2022). Digital health interventions for weight management in children and adolescents: Systematic review and meta-analysis. *Journal of Medical Internet Research, 24*(2), Article e30675. https://doi.org/10.2196/30675

Li, L., Sun, F., Du, J., Li, Z., Chen, T., & Shi, X. (2025). Behavior-change lifestyle interventions for the treatment of obesity in children and adolescents: A scoping review. *Annals of the New York Academy of Sciences, 1543*(1), 31–41. https://doi.org/10.1111/nyas.15278

Melo, G. L. R., Santo, R. E., Mas Clavel, E., Bosque Prous, M., Koehler, K., Vidal-Alaball, J., van der Waerden, J., Gobiņa, I., López-Gil, J. F., Lima, R., & Agostinis-Sobrinho, C. (2025). Digital dietary interventions for healthy adolescents: A systematic review of behavior change techniques, engagement strategies, and adherence. *Clinical Nutrition, 45*, 176–192. https://doi.org/10.1016/j.clnu.2025.01.012

Metzendorf, M.-I., Wieland, L. S., & Richter, B. (2024). Mobile health (m-health) smartphone interventions for adolescents and adults with overweight or obesity. *Cochrane Database of Systematic Reviews, 2024*(2), Article CD013591. https://doi.org/10.1002/14651858.CD013591.pub2

Milne-Ives, M., Homer, S., Andrade, J., & Meinert, E. (2024). The conceptualisation and measurement of engagement in digital health. *Internet Interventions, 36*, Article 100735. https://doi.org/10.1016/j.invent.2024.100735

Ministry of Health of the Republic of Indonesia. (2023). *Survei Kesehatan Indonesia (SKI) 2023 dalam angka [2023 Indonesian Health Survey in figures]*. https://www.badankebijakan.kemkes.go.id/ski-2023-dalam-angka/

NCD Risk Factor Collaboration (NCD-RisC). (2024). Worldwide trends in underweight and obesity from 1990 to 2022: A pooled analysis of 3,663 population-representative studies with 222 million children, adolescents, and adults. *The Lancet, 403*(10431), 1027–1050. https://doi.org/10.1016/S0140-6736(23)02750-2

Oh, Y. J., Zhang, J., Fang, M.-L., & Fukuoka, Y. (2021). A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. *International Journal of Behavioral Nutrition and Physical Activity, 18*(1), Article 160. https://doi.org/10.1186/s12966-021-01224-6

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

Perski, O., Blandford, A., West, R., & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine, 7*(2), 254–267. https://doi.org/10.1007/s13142-016-0453-1

Prafiantini, E., Agustina, R., Purwandari, B., Chandra, D. N., Bintari, D. R., Rajwadini, F. Z., Farhanah, J., & Hendarto, A. (2025). Usability testing of EatsUp®: Mobile application for monitoring balanced dietary practices and active lifestyle among adolescents—A study in Jakarta, Indonesia. *Frontiers in Digital Health, 7*, Article 1506952. https://doi.org/10.3389/fdgth.2025.1506952

Putri, D. S. K., Sari, K., Utami, N. H., Nazarina, N., Amelia, T., Salwa, N. Y., Sulistiyowati, N., Ega, A. V., Azzumar, M., Rachmawati, R., Salimar, Despitasari, M., & Mulyantoro, D. K. (2025). Usability evaluation of the "Teen 'n Fit" mobile health application: A formative study among Indonesian adolescent girls. *PLOS ONE, 20*(11), Article e0337013. https://doi.org/10.1371/journal.pone.0337013

Rachmi, C. N., Jusril, H., Ariawan, I., Beal, T., & Sutrisna, A. (2021). Eating behaviour of Indonesian adolescents: A systematic review of the literature. *Public Health Nutrition, 24*(S2), S84–S97. https://doi.org/10.1017/S1368980020002876

Rahayu, S. R., Sulistiani, D., Yuniastuti, A., & Merzistya, A. N. A. (2021). "Gizi COVID-19 bagi remaja" application as a COVID-19 prevention. *KEMAS: Jurnal Kesehatan Masyarakat, 17*(1), 109–119. https://doi.org/10.15294/kemas.v17i1.26843

Ravelli, M. N., & Schoeller, D. A. (2020). Traditional self-reported dietary instruments are prone to inaccuracies and new approaches are needed. *Frontiers in Nutrition, 7*, Article 90. https://doi.org/10.3389/fnut.2020.00090

Roshita, A., Riddell-Carre, P., Sjahrial, R., Jupp, D., Torlesse, H., Izwardy, D., & Rah, J. H. (2021). A qualitative inquiry into the eating behavior and physical activity of adolescent girls and boys in Indonesia. *Food and Nutrition Bulletin, 42*(1_suppl), S122–S131. https://doi.org/10.1177/0379572121990948

Sardi, L., Idri, A., & Fernández-Alemán, J. L. (2017). A systematic review of gamification in e-health. *Journal of Biomedical Informatics, 71*, 31–48. https://doi.org/10.1016/j.jbi.2017.05.011

Shatwan, I. M., Alhefani, R. S., Bukhari, M. F., Hanbazazah, D. A., Srour, J. K., Surendran, S., Aljefree, N. M., & Almoraie, N. M. (2023). Effects of a smartphone app on fruit and vegetable consumption among Saudi adolescents: Randomized controlled trial. *JMIR Pediatrics and Parenting, 6*, Article e43160. https://doi.org/10.2196/43160

Simmonds, M., Llewellyn, A., Owen, C. G., & Woolacott, N. (2016). Predicting adult obesity from childhood obesity: A systematic review and meta-analysis. *Obesity Reviews, 17*(2), 95–107. https://doi.org/10.1111/obr.12334

Stasinaki, A., Büchter, D., Shih, C.-H. I., Heldt, K., Güsewell, S., Brogle, B., Farpour-Lambert, N., Kowatsch, T., & l'Allemand, D. (2021). Effects of a novel mobile health intervention compared with a multicomponent behaviour-changing program on body mass index, physical capacities, and stress parameters in adolescents with obesity: A randomized controlled trial. *BMC Pediatrics, 21*(1), Article 308. https://doi.org/10.1186/s12887-021-02781-2

Stephens, T. N., Joerin, A., Rauws, M., & Werk, L. N. (2019). Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. *Translational Behavioral Medicine, 9*(3), 440–447. https://doi.org/10.1093/tbm/ibz043

Suleiman-Martos, N., García-Lara, R. A., Martos-Cabrera, M. B., Albendín-García, L., Romero-Béjar, J. L., Cañadas-De la Fuente, G. A., & Gómez-Urquiza, J. L. (2021). Gamification for the improvement of diet, nutritional habits, and body composition in children and adolescents: A systematic review and meta-analysis. *Nutrients, 13*(7), Article 2478. https://doi.org/10.3390/nu13072478

Thompson, D., & Baranowski, T. (2019).

Chatbots as extenders of pediatric obesity intervention: An invited commentary on "Feasibility of pediatric obesity and pre-diabetes treatment support through Tess, the AI behavioral coaching chatbot." *Translational Behavioral Medicine, 9*(3), 448–450. https://doi.org/10.1093/tbm/ibz065

Tugault-Lafleur, C. N., De-Jongh González, O., Macdonald, J., Bradbury, J., Warshawski, T., Ball, G. D. C., Morrison, K., Ho, J., Hamilton, J., Buchholz, A., & Mâsse, L. (2023). Efficacy of the Aim2Be intervention in changing lifestyle behaviors among adolescents with overweight and obesity: Randomized controlled trial. *Journal of Medical Internet Research, 25*, Article e38545. https://doi.org/10.2196/38545

Zhang, X., Liu, J., Ni, Y., Yi, C., Fang, Y., Ning, Q., Shen, B., Zhang, K., Liu, Y., Yang, L., Li, K., Liu, Y., Huang, R., & Li, Z. (2024). Global prevalence of overweight and obesity in children and adolescents: A systematic review and meta-analysis. *JAMA Pediatrics, 178*(8), 800–813. https://doi.org/10.1001/jamapediatrics.2024.1576