South American Publishing

SAP Gamification and Augmented Reality. 2023;2:13
doi: 10.62486/gr202313

ORIGINAL

Gamification mechanics in top-rated educational apps by audience age

Mecánicas de gamificación en aplicaciones educativas mejor valoradas según la edad del público

Lucía Gerez[1]
Verónica Daniela Larrosa Valls[2]

[1] Universidad Nacional de Tucumán, Tucumán, Argentina
[2] Universidad de Buenos Aires, Buenos Aires, Argentina

Submitted: 08-03-2023 | Revised: 05-05-2023 | Accepted: 09-08-2023 | Published: 11-08-2023

Corresponding Author: Lucía Gerez

ABSTRACT

Introduction: Gamification is widely proposed to sustain motivation in digital learning, yet it is unclear which mechanics highly rated educational apps advertise and whether they differ by the age of the intended users. This study estimated the disclosure of points, badges, leaderboards and streaks in top-rated educational apps and compared it across target age groups. Methods: Cross-sectional content analysis of public App Store metadata. Education chart entries from 14 storefronts were restricted to apps first released in 2010-2023 and rated 4.5 or higher with at least 1000 ratings. Mechanics were detected with multilingual keyword dictionaries applied to names and descriptions, and audience was classified as early childhood, school age or general or adult. Wilson intervals, chi-square or permutation tests with Holm adjustment, and logistic regression were used. Results: Of 964 charted apps, 243 were top-rated. At least one core mechanic was disclosed by 28 apps (11.5%); leaderboards (4.9%), streaks (4.5%), points (2.1%) and badges (2.1%) were uncommon, and none differed by audience after adjustment. Rewards or virtual currency were disclosed by 35.3% of early childhood apps versus 2.8% of school age apps and 3.2% of general or adult apps (adjusted p = 0.006). Conclusions: Listings of top-rated educational apps rarely name the four classic mechanics, and age group changes little beyond a concentration of rewards in early childhood listings. These results describe advertised features in one store and require in-app validation.

Keywords: gamification; mobile applications; educational technology; game design elements; app stores.

RESUMEN

Introducción: La gamificación se propone para sostener la motivación en el aprendizaje digital, pero se desconoce qué mecánicas anuncian las aplicaciones educativas mejor valoradas y si varían con la edad. Se estimó la declaración de puntos, insignias, clasificaciones y rachas y se comparó entre grupos de edad. Métodos: Análisis de contenido transversal de metadatos públicos de la App Store. Las listas de Educación de 14 tiendas se restringieron a aplicaciones lanzadas en 2010-2023 con valoración de 4,5 o más y al menos 1000 valoraciones. Las mecánicas se detectaron con diccionarios multilingües en nombres y descripciones, y el público se clasificó como primera infancia, edad escolar o general o adulto. Se usaron intervalos de Wilson, pruebas de asociación con ajuste de Holm y regresión logística. Resultados: De 964 aplicaciones en las listas, 243 fueron de alta valoración. Al menos una mecánica central se declaró en 28 (11,5%); tablas de clasificación (4,9%), rachas (4,5%), puntos (2,1%) e insignias (2,1%) fueron poco frecuentes y ninguna difirió según el público. Las recompensas o monedas virtuales se declararon en 35,3% de las aplicaciones de primera infancia frente a 2,8% de edad escolar y 3,2% de público general o adulto (p ajustada = 0,006). Conclusiones: Las fichas rara vez nombran las cuatro mecánicas clásicas y la edad cambia poco, salvo por más recompensas en la primera infancia. Los resultados describen funciones anunciadas en una tienda y requieren validación interna.

Palabras clave: gamificación; aplicaciones móviles; tecnología educativa; elementos de diseño de juegos; tiendas de aplicaciones.

© 2023; The Authors. This Open access article is distributed under CC BY 4.0 licence.

INTRODUCTION

Gamification, the use of game elements in activities that are not games, has become a default response to the problem of sustaining learner motivation in digital settings.(1,2) Its research base is large but fragmented: a systematic review of the theoretical foundations of gamification, serious games and game-based learning identified 118 different theories,(3)and recent meta-analyses report mixed effects on learning outcomes.(4,5) Negative effects have been documented as well, for example lower content knowledge and course satisfaction when assessment activities were gamified,(6)and a systematic mapping has examined negative effects in educational software.(7) Part of this heterogeneity may come from the absence of a shared classification of gamification elements, which makes studies difficult to compare.(1,8) Interest has grown quickly, as bibliometric work shows,(9)and reviews now cover science education,(10)the motivation of high school and university students(11)and e-learning in higher education, where low completion rates have been attributed to a lack of student motivation.(12)Studies that sample apps directly from public stores offer an alternative view, because they describe what users are offered rather than what a trial tested. In health and fitness, gamification was present in 64% of the 50 top-ranked free App Store applications, but points and levels appeared in only 6% and 3% of them;(13)stress management apps from Google Play used an average of 0.5 gamification techniques;(14)and a content analysis of 754 gamification cases found that points, leaderboards and progress dominated the collection.(15)Educational apps have received less attention of this kind, although the App Store hosts an abundance of titles labelled as educational whose claims are often unsubstantiated.(16)

The evidence on educational apps comes mostly from single-app evaluations and reviews. Duolingo was used in third and fourth grade Spanish instruction,(17)two popular English vocabulary apps were analysed through a mechanics, dynamics and aesthetics framework,(18)a kindergarten vocabulary app was built around game mechanics,(19)and reviews cover mobile-assisted and gamification-based language learning.(20) For primary school numeracy, levels of difficulty and feedback from the current level made a game version more effective and engaging than a version without them,(2)whereas among distance university students different gamification user types were related to different online learning activities.(21)These findings suggest that useful mechanics may depend on the age of the learner: it is plausible that young children are served by immediate feedback and reward, whereas adult learners may respond to progress paths and social comparison, which gamification theory describes as separate functions of game elements.(3) Top-rated apps deserve separate attention, because ratings and reviews influence downloads(22) and the apps with the best ratings are therefore the ones most likely to shape how learners meet gamification. Even so, no study has compared the mechanics advertised by top-rated educational apps across target age groups, and it is unknown whether the four classic mechanics (points, badges, leaderboards and streaks) are in fact the ones that distinguish the best-rated products. This study aimed to estimate the disclosure of points, badges, leaderboards and streaks, together with five secondary mechanics, in top-rated educational apps of the App Store, and to compare it across early childhood, school age and general or adult audiences.

METHODS

A cross-sectional content analysis of public store metadata was conducted. The unit of analysis was the unique app, and the outcome was the disclosure of a gamification mechanic in the app name or store description. The analysis was descriptive and exploratory, and the items of the STROBE statement that apply to cross-sectional designs were followed.(23)No human participants were involved.

Chart entries of the Education genre (genre identifier 6017) were obtained from the public iTunes RSS feeds of the free and paid charts in 14 App Store storefronts: the United States, United Kingdom, Canada, Australia, Mexico, Brazil, Argentina, Colombia, Chile, Peru, Spain, Germany, France and Italy. Each chart supplied up to 100 entries, and the six Latin American storefronts were included to allow a discussion of transferability to that region. Metadata for every chart entry (name, description, price, average rating, number of ratings, release date, genres and store age rating) were then retrieved from the iTunes Lookup API of each storefront.(24)All requests were automatic and required no login, and the data correspond to the snapshot available from the public API. Google Play was not used because its listings are not exposed through an open programmatic interface without login, so every conclusion refers to the App Store. The requests were spaced in time to respect the service limits, and each lookup request carried up to 100 app identifiers. Listing text

is supplied by the developer, so it reflects what the developer chooses to advertise. Chart-based sampling follows earlier work on top-ranked apps,(13,25) and store descriptors have been used to characterise apps in other fields.(26,27)

Apps were eligible when Education was listed among their genres and the year of first release fell between 2010 and 2023. Apps that appeared in several storefronts were counted once, and the record with the largest number of ratings was retained, because ratings are storefront specific. Top-rated apps were those with an average rating of 4.5 or higher and at least 1000 ratings; both thresholds were fixed in the protocol before any mechanic was coded. Charted apps without ratings were therefore classed as not top-rated. Because presence in a chart depends on seller and app attributes(25)whereas ratings are shaped by app store strategy,(28)charted apps below the thresholds were kept as a comparison group for a sensitivity analysis.

The average rating and the number of ratings were taken from the retained storefront record, price was the list price in the storefront currency (free when zero), and the release year was the year of the first release date. The store age rating, the list of genres and the storefront of the retained record were also kept. The Game Center flag, which marks apps connected to Apple's native service for leaderboards and achievements, was used as a structured indicator that does not depend on listing text. Store descriptions had a median length of 2,397 characters (P25-P75, 1,452-3,453), none was empty, and each description is written in the language of the storefront that supplied the retained record, with 72 apps retained from the United States storefront, 33 from Germany and 30 from Brazil.

Three mutually exclusive target audiences were assigned in a fixed order of precedence. Apps whose name or first 600 characters of description contained terms for young children (for example kids, toddlers, preschool, kindergarten or their Spanish, Portuguese, French, German and Italian equivalents) were classed as early childhood. Among the remaining apps, those with terms for formal schooling (students, grade, school, classroom, curriculum, teachers and equivalents) were classed as school age, and all others as general or adult. The store age rating (4+, 9+, 12+ and 17+) was kept as a descriptive variable, with 12+ and 17+ combined because of the small numbers. Audience classification is therefore based on the stated target of the listing and not on the age rating.

Nine mechanics were defined from published taxonomies of game elements.(1,15) Four were primary: points, badges or achievements, leaderboards or rankings, and streaks. Five were secondary: levels, rewards or virtual currency, challenges or missions, progress tracking and avatars. Each mechanic was detected with regular expressions applied to the app name and description in the six languages of the storefronts. Patterns required a context of action, such as earning points, unlocking levels or keeping a daily streak, so that ordinary uses of words like point, level, mission or challenge were not counted. The dictionaries were refined once after reading the matched text segments of the analytic sample to remove false positives, and they were not validated against in-app inspection or against independent raters. No artificial intelligence system was used to code the mechanics. The outcome is therefore the disclosure of a mechanic in listing text, and the absence of a mention does not prove that the mechanic is absent from the app.

Several sources of bias were anticipated. Chart sampling retains popular apps and misses highly rated apps that never reached a chart, so the sample describes the visible top of the market and not all top-rated apps. Keyword detection can misclassify a listing in either direction, and the effect is likely to differ by language because the dictionaries were written first for English. Audience misclassification is possible when a listing mentions several groups, in which case the order of precedence assigns the youngest. Missing data were handled by exclusion: apps without ratings were not imputed and were not treated as top-rated, and no analytic variable had missing values among the top-rated apps.

No sample size calculation was performed, because the study analysed every top-rated app found in the charts of the selected storefronts, and the precision of each estimate is conveyed by its confidence interval. The number of early childhood apps was fixed by the data, so comparisons that involve this group have limited power and are interpreted with the effect sizes and intervals in view, not with significance tests alone.

The dictionaries combined patterns in English, Spanish, Portuguese, French, German and Italian. Points were matched when a listing referred to earning, collecting or redeeming points or experience points, or to a points system. Badges or achievements required the word badge, trophy or medal, or an action verb such as earn or unlock together with achievements. Leaderboards or rankings covered leaderboard, league, weekly or global ranking, high score and competing with other users. Streaks covered the word streak, consecutive days and days in a row, and their equivalents in the other languages. Levels required an action context such as unlocking levels, rewards covered rewards, prizes, coins and gems, and challenges or missions covered daily or weekly challenges,

quests and missions. Progress tracking covered progress bars, progress reports and phrases about tracking progress, and avatars covered avatars and character customization. Negated mentions, such as a listing stating that a game has no high scores, were excluded.

A protocol was written before the analysis. It specified the research question, the unit of analysis, the eligibility window, the top-rated thresholds, the definition of audience, the nine mechanics, the tests, the multiplicity correction and the sensitivity analyses. Amendments are limited to the regression outcome described below, and decisions taken after inspection of the listing text are reported in this section.

Proportions were reported with Wilson 95% confidence intervals. The association between audience and each mechanic was tested with the chi-square test, replaced by a Monte Carlo permutation test with 3000 permutations and a fixed seed when any expected count was below 5, and p values were adjusted across the nine mechanics with the Holm method. Cramér V was reported as the effect size. The number of mechanics per app was compared across audiences with the Kruskal-Wallis test. The prespecified model was a logistic regression for the disclosure of at least three of the nine mechanics, but only 6 apps met that outcome and the estimates were unstable, so after seeing the descriptive counts the outcome was changed to the disclosure of at least one of the four core mechanics, with audience (school age as reference), price, release period (2010-2016 versus 2017-2023) and the base-10 logarithm of the number of ratings as covariates. This change is a post hoc amendment. Two sensitivity analyses were run: restriction to the four English-language storefronts, in which the dictionaries are most reliable, and comparison of the top-rated apps with the remaining charted apps that had ratings. Tests were two-sided with a significance level of 0.05, and all analyses were exploratory.

The study was descriptive. No hypothesis about the direction of differences was stated in advance, and every test is reported, including those that did not reach significance, so that readers can judge how much of the evidence is null. Percentages use the number of apps of each audience as the denominator, and the denominator for the overall prevalences is the 243 top-rated apps. Tables report counts and percentages together, and the figure shows the seven mechanics with the highest disclosure.

RESULTS

The chart feeds returned 2,774 records from 14 storefronts, which collapsed to 1,403 unique educational apps, and 439 of them were first released before 2010 or after 2023. Of the 964 charted apps that remained, 243 (25.2%) met the top-rated criterion, while 84 had no ratings. The analytic sample comprised 154 general or adult apps (63.4%), 72 school age apps (29.6%) and 17 early childhood apps (7.0%). The median rating was 4.73, and the median number of ratings amounted to 15,613 (P25-P75, 3,870-47,951). Free apps were 211 (86.8%), 136 (56.0%) were first released in 2017-2023 and 196 (80.7%) carried a store age rating of 4+, a distribution that did not differ by audience (p = 0.104). Only 1 app (0.4%) had Game Center enabled, and 102 (42.0%) were listed in at least one Latin American storefront (Table 1).

Table 1. Characteristics of top-rated educational apps by target audience.

Characteristic	Early childhood	School age	General or adult	All
Apps, n	17	72	154	243
Rating, median	4.71	4.73	4.73	4.73
Rating count, median (P25-P75)	37202 (12300-78279)	28492 (6909-85319)	9584 (3301-37719)	15613 (3870-47951)
Free of charge, n (%)	16 (94.1)	66 (91.7)	129 (83.8)	211 (86.8)
First released 2017-2023, n (%)	12 (70.6)	35 (48.6)	89 (57.8)	136 (56.0)
Store age rating 4+, n (%)	17 (100.0)	62 (86.1)	117 (76.0)	196 (80.7)
Store age rating 9+, n (%)	0 (0.0)	1 (1.4)	5 (3.2)	6 (2.5)
Store age rating 12+ or 17+, n (%)	0 (0.0)	9 (12.5)	32 (20.8)	41 (16.9)
Game Center enabled, n (%)	0 (0.0)	0 (0.0)	1 (0.6)	1 (0.4)

Note: Top-rated is an average rating of 4.5 or higher with at least 1000 ratings. P25-P75 = interquartile range. Audience was assigned from the app name and the first 600 characters of the store description.

At least one core mechanic was disclosed by 28 apps (11.5%), two or more by 5 (2.1%), and none disclosed all four. Leaderboards or rankings were the most frequent core mechanic (12; 4.9%, 95% CI 2.8-8.4), followed by streaks (11; 4.5%, 95% CI 2.5-7.9), points (5; 2.1%) and badges or achievements (5; 2.1%). The proportion with at least one core mechanic was 11.8% in early childhood, 8.3% in school age and 13.0% in general or adult apps, which did not differ (p = 0.594). No single core mechanic was associated with audience after Holm adjustment, the smallest adjusted p value being 0.296 for badges. The median number of mechanics was zero in every audience (p = 0.163), and the store age rating did not separate apps either (p = 0.394). In turn, leaderboards were disclosed by 10 general or adult apps and 2 school age apps, and by no early childhood app (Table 2).

Table 2. Gamification mechanics disclosed in store listings by target audience.

Mechanic	Early childhood	School age	General or adult	All, n (%; 95% CI)	p (Holm)	Cramér V
Points	0 (0.0)	1 (1.4)	4 (2.6)	5 (2.1; 0.9-4.7)	1.000	0.06
Badges or achievements	2 (11.8)	1 (1.4)	2 (1.3)	5 (2.1; 0.9-4.7)	0.296	0.19
Leaderboards or rankings	0 (0.0)	2 (2.8)	10 (6.5)	12 (4.9; 2.8-8.4)	1.000	0.10
Streaks	0 (0.0)	3 (4.2)	8 (5.2)	11 (4.5; 2.5-7.9)	1.000	0.06
Levels	0 (0.0)	0 (0.0)	2 (1.3)	2 (0.8; 0.2-3.0)	1.000	0.07
Rewards or virtual currency	6 (35.3)	2 (2.8)	5 (3.2)	13 (5.3; 3.2-8.9)	0.006	0.37
Challenges or missions	0 (0.0)	1 (1.4)	1 (0.6)	2 (0.8; 0.2-3.0)	1.000	0.04
Progress tracking	2 (11.8)	8 (11.1)	24 (15.6)	34 (14.0; 10.2-18.9)	1.000	0.06
Avatars	2 (11.8)	0 (0.0)	1 (0.6)	3 (1.2; 0.4-3.6)	0.099	0.26

Note: Values are n (%) of apps whose listing text matched the keyword dictionary for the mechanic. CI = confidence interval (Wilson). The p values come from chi-square tests, or from permutation tests when expected counts were below 5, and were adjusted with the Holm method across nine mechanics. Cramér V is the effect size.

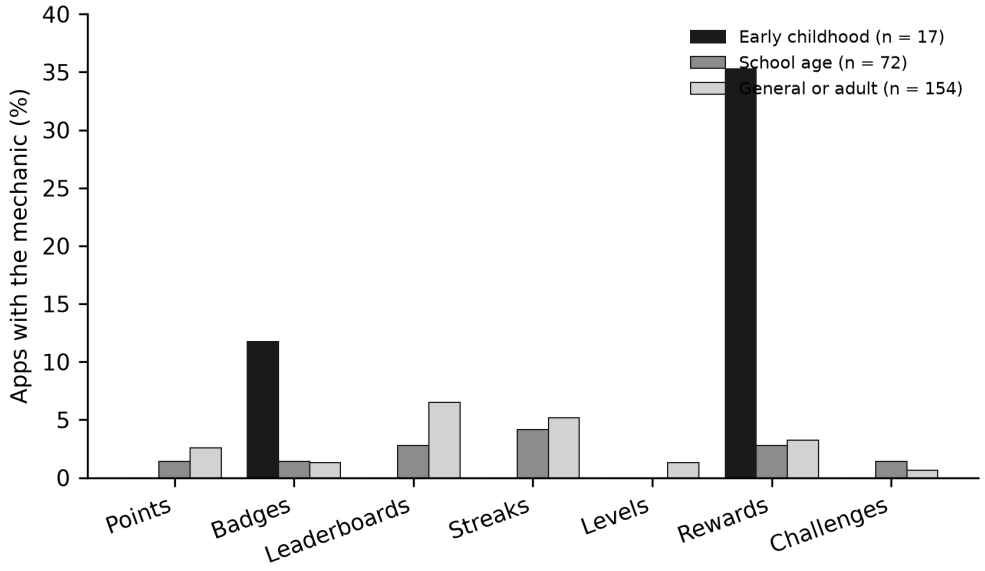
Rewards or virtual currency were the only mechanic that varied with audience: 6 of 17 early childhood apps (35.3%) disclosed them, against 2.8% of school age and 3.2% of general or adult apps (adjusted p = 0.006; Cramér V = 0.37). Avatars were disclosed by 2 early childhood apps and by one other app, a difference that did not survive adjustment (adjusted p = 0.099). Progress tracking was the most frequent secondary mechanic (34; 14.0%), whereas levels (2) and challenges or missions (2) were rare, and only 5 listings used the word gamification. In the adjusted model, a first release in 2017-2023 was associated with the disclosure of a core mechanic (OR = 2.56; 95% CI 0.99-6.67; p = 0.053), but audience was not (early childhood OR = 1.13; general or adult OR = 1.86) (Table 3). Instead, the 637 remaining charted apps with ratings disclosed a core mechanic at a lower proportion than the top-rated apps (5.5% versus 11.5%), with streaks differing most (p <0.001). In the 117 apps retained from English-language storefronts, 12.0% disclosed a core mechanic and audience again made no difference (p = 0.544), and apps listed in a Latin American storefront did not differ from the rest (11.8% versus 11.3%; p = 0.920) (Figure 1).

Table 3. Adjusted odds of disclosing at least one core mechanic.

Variable	Adjusted OR	95% CI	p
Early childhood (vs school age)	1.13	0.20-6.34	0.888
General or adult (vs school age)	1.86	0.69-5.03	0.221
Paid (vs free)	0.40	0.05-3.31	0.393
First released 2017-2023 (vs 2010-2016)	2.56	0.99-6.67	0.053
Rating count (per tenfold increase)	1.58	0.92-2.71	0.099

Note: Logistic regression with 243 apps. Core mechanics are points, badges, leaderboards and streaks. OR = odds ratio; CI = confidence interval.

Figure 1. Disclosure of gamification mechanics in top-rated educational apps by target audience.



Note: Bars show the percentage of apps in each audience whose listing text matched the dictionary of each mechanic. The number of apps in each audience is given in the legend.

DISCUSSION

The listings of top-rated educational apps seldom name the four classic mechanics, and the age of the intended audience explains little of what they do name. This indicates that high consumer ratings in education are not tied to an advertised points, badges, leaderboards and streaks package, and that the language used to sell an educational app is closer to feedback and progress than to competition, an observation that matters for the transfer of gamification design to Latin American classrooms, where 42.0% of the top-rated apps were listed in at least one regional storefront. The pattern fits the theoretical view that gamification works through goals, immediate feedback and guided paths, with social comparison as only one of several options,(3)and it agrees with the observation that gamification depends on context more than on a fixed set of elements.(8)One reason may be that a listing is written to attract downloads and that the educational label of many store apps is unsubstantiated,(16)so the vocabulary of a listing is an imperfect window onto design. The one mechanic that varied, rewards or virtual currency, was concentrated in listings aimed at young children, and its effect size was moderate to large, which suggests that the audience effect, where it exists, acts through the type of reinforcement and not through the number of mechanics. The negative findings for the four core mechanics should not be read as proof of equality across audiences, because the early childhood group was small: badges, for instance, were disclosed by 11.8% of early childhood apps and 1.3% of general or adult apps, a gap that did not survive adjustment but whose interval is wide. The same reading applies to the secondary mechanics: progress tracking was the most frequent, and it overlaps with the feedback and progress principles that the theoretical literature places at the core of gamification design.(3)

Several earlier app-store studies agree with the low disclosure observed here. In the 50 top-ranked free health and fitness apps, points and levels were uncommon, and the elements most often present were goal setting, social influences and challenges.(13)Stress management apps used an average of 0.5 gamification techniques, and the use of techniques was unrelated to behavior change techniques.(14)Among the most downloaded ecology apps that did use game elements, levels, achievements, challenges and feedback predominated, with narratives and emotions as components of the experience.(29)The concentration of rewards in early childhood listings also matches design reports for young children: a letter writing app rewarded successful handwriting with an animation,(16)children identified levels of difficulty and feedback as appealing characteristics of games,(2)virtual characters were among the most frequent features of children's oral hygiene apps,(30)and gamified diabetes apps built around caring for a character used a similar logic.(27)In addition, free apps made up most of the sample, in line with the longer survival of free apps in top charts,(25)which means that the educational apps seen at the top are those whose

business model favors broad adoption over paid features. Gamification and game-like images are also used as engagement strategies in nutrition apps intended for children,(31)and the avatars disclosed almost exclusively by early childhood apps, although not significant after adjustment, point in the same direction.

Other studies reached different proportions, and the differences can be explained by sampling and measurement. A content analysis of 754 gamification cases found that points, leaderboards and progress accounted for a large share of the collection,(15)and apps selected because they were gamified averaged 6.87 of 31 gamification features in children's oral hygiene(30)or were grouped into archetypes in a database of 143 health apps.(32)Those samples were built from apps already identified as gamified, whereas the present sample was defined by consumer ratings in the Education chart, so a lower prevalence is expected. Studies of language apps also describe richer mechanics than the listings suggest: two popular English vocabulary apps were more useful to learners when they rewarded them with in-kind incentives and added social interaction,(18)and classroom gamification for school-age learners can be designed by teachers around the psychological needs of students and not embedded in the app that they use.(33)If that holds for school-age audiences, it might explain why the listings of apps used in classrooms disclose few mechanics even when the teaching around them is gamified. A further source of difference is the effect of rewards themselves: gamification can have unintended effects when assessment is gamified,(6) extrinsic rewards need to be balanced against intrinsic motivation,(34)and practitioners report negative effects in educational software,(7)so developers who target children or students may avoid advertising competitive elements. Taxonomies also differ in what counts as a mechanic,(1)and the earlier health studies counted elements such as goal setting and social influences that were outside the nine mechanics used here, so proportions across studies are not directly comparable. The slightly higher share of leaderboards among general or adult apps than among apps for younger audiences, which was descriptive and not significant, is compatible with the heterogeneity of gamification user types reported for adult distance learners,(21)but the present data cannot test that account. These explanations are plausible but were not tested here. The higher disclosure of core mechanics among top-rated apps than among other charted apps also needs a careful reading, because ratings and reviews influence downloads and app store strategy influences ratings,(22,28) and because quality reviews of mental health apps have prioritized clinician and academic views over those of users.(35)A rating therefore reflects consumer satisfaction and not evidence of educational effectiveness, which a content analysis of listings cannot assess.

Several limitations qualify these results, and each suggests a specific next step. First, the outcome is the mention of a mechanic in listing text, not its verified presence or use, so prevalences are probably underestimated and may be biased toward apps with long descriptions; future studies should code mechanics inside the apps with two trained raters and report agreement, as was done for stress management apps.(14) Second, the dictionaries were refined after inspecting the matched text and were not validated, and the six languages are covered unevenly, so a validation sample coded by raters and a language-specific sensitivity analysis should accompany any replication. Third, only the App Store was analysed because Google Play offers no open programmatic interface, so the results cannot be generalized across app stores, and the apps that dominate one store may not dominate the other; studies with access to Play data should replicate the design, as earlier reviews that searched several stores did.(26)Fourth, audience was inferred from the wording of listings, and the store age rating barely varied, which limits the comparison; developer-declared target ages or expert coding of the content would give a sounder exposure variable, and a second classification could be compared with the first to quantify misclassification. Fifth, the early childhood group contained only 17 apps, so differences smaller than those found for rewards could not be detected, and the logistic model had wide intervals; oversampling young-child apps is needed, together with confidence intervals that make the imprecision visible, as was done here for the overall prevalences. Sixth, the ratings and counts reflect the snapshot available from the public API, which includes ratings accumulated after 2023 for apps released earlier, so versioned data would allow analysis by period. Seventh, each chart supplied only 100 entries per storefront, so apps that rank lower were never observed; sampling by search across the Education genre and weighting by downloads, where available, would show whether the pattern holds below the top charts. Eighth, apps listed in several storefronts were counted once and carried the ratings of a single storefront, and 95 of the retained apps (39.1%) were listed in more than one, so storefront-level analyses with ratings kept separate would show whether the pattern differs by market. Ninth, the outcome of the regression was changed after the descriptive counts were seen, and all tests were exploratory, so the associations reported should be treated as hypotheses.

For educators and families, the results imply that the listing of a highly rated educational

app is a weak guide to the motivational design inside it, and that reading a description for points, badges or leaderboards would miss most of what the app does. For developers, the dominance of progress tracking and the concentration of rewards in early childhood listings point to a design space in which feedback is advertised more readily than competition, which is consistent with the concern about narrow or extrinsic gamification.(6,34) Listings in Spanish and Portuguese make up most of the Latin American material, and the keyword patterns were not checked separately by language, so detection may be less sensitive for these listings than for English ones. Only 59 of the retained records came from the six Latin American storefronts, and 102 apps were listed in at least one of them, so the regional comparison is limited by its size. For readers in Latin America, apps listed in the six Latin American storefronts disclosed core mechanics at a similar rate to the rest, which supports transferability of the pattern, although the Spanish and Portuguese dictionaries were less tested than the English ones. Dictionaries built on published taxonomies could be shared and extended to other languages and stores, which would allow the disclosure of mechanics to be tracked over time and compared across markets. Researchers who work with store data should treat listing text as a measure of advertised design and not as a measure of use, and should link it to in-app coding and to learning outcomes, which the heterogeneity of effects in meta-analyses shows is still unsettled.(4,5) Reviewers of educational apps could add a check of advertised mechanics against in-app behavior to the quality frameworks they already use, so that a listing claiming a leaderboard or a streak can be verified and so that apps whose engagement relies on rewards for young children can be examined for the balance between extrinsic and intrinsic motivation.

CONCLUSIONS

Points, badges, leaderboards and streaks were rarely disclosed in the listings of top-rated educational apps, and their disclosure did not differ by target age group. Rewards or virtual currency were the only mechanic that varied, concentrated in early childhood listings, while progress tracking was the most frequent secondary mechanic. These findings describe advertised features in one store and require validation inside the apps before they guide design or selection.

REFERENCES

1. Schöbel SM, Janson A, Söllner M. Capturing the complexity of gamification elements: a holistic approach for analysing existing and deriving novel gamification designs. European Journal of Information Systems. 2020;29(6):641-668. doi: 10.1080/0960085x.2020.1796531
2. Nand K, Baghaei N, Casey J, Barmada B, Mehdipour F, Liang HN. Engaging children with educational content via Gamification. Smart Learn Environ. 2019;6(1):6. doi: 10.1186/s40561-019-0085-2
3. Krath J, Schürmann L, von Korflesch HFO. Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. Computers in Human Behavior. 2021;125:106963. doi: 10.1016/j.chb.2021.106963
4. Li M, Ma S, Shi Y. Examining the effectiveness of gamification as a tool promoting teaching and learning in educational settings: a meta-analysis. Front Psychol. 2023;14:1253549. doi: 10.3389/fpsyg.2023.1253549
5. Costa JM. Using game concepts to improve programming learning: A multi-level meta-analysis. Comp Applic In Engineering. 2023;31(4):1098-1110. doi: 10.1002/cae.22630
6. Kwon HY, Özpolat K. The Dark Side of Narrow Gamification: Negative Impact of Assessment Gamification on Student Perceptions and Content Knowledge. INFORMS Transactions on Education. 2021;21(2):67-81. doi: 10.1287/ited.2019.0227
7. Almeida C, Kalinowski M, Uchôa A, Feijó B. Negative effects of gamification in education software: Systematic mapping and practitioner perceptions. Information and Software Technology. 2023;156:107142. doi: 10.1016/j.infsof.2022.107142
8. Helmefalk M. An interdisciplinary perspective on gamification: Mechanics, psychological mediators and outcomes. IJSG. 2019;6(1):3-26. doi: 10.17083/ijsg.v6i1.262
9. Trinidad M, Ruiz M, Calderon A. A Bibliometric Analysis of Gamification Research. IEEE Access. 2021;9:46505-46544. doi: 10.1109/access.2021.3063986
10. Kalogiannakis M, Papadakis S, Zourmpakis AI. Gamification in Science Education. A Systematic Review of the Literature. Education Sciences. 2021;11(1):22. doi: 10.3390/educsci11010022
11. Ratinho E, Martins C. The role of gamified learning strategies in student's motivation in high school and higher education: A systematic review. Heliyon. 2023;9(8):e19033. doi: 10.1016/j.heliyon.2023.e19033
12. Khaldi A, Bouzidi R, Nader F. Gamification of e-learning in higher education: a systematic literature review. Smart Learn Environ. 2023;10(1):10. doi: 10.1186/s40561-023-00227-z
13. Cotton V, Patel MS. Gamification Use and Design in Popular Health and Fitness Mobile Applications. Am J Health Promot. 2019;33(3):448-451. doi: 10.1177/0890117118790394
14. Hoffmann A, Christmann CA, Bleser G. Gamification in Stress Management Apps: A Critical App Review. JMIR Serious Games. 2017;5(2):e13. doi: 10.2196/games.7216
15. Park S, Kim S. Patterns Among 754 Gamification Cases: Content Analysis for Gamification Development. JMIR Serious Games. 2018;6(4):e11336. doi: 10.2196/11336
16. Martens M, Rinnert GC, Andersen C. Child-Centered Design: Developing an Inclusive Letter Writing App. Front Psychol. 2018;9(DEC):2277. doi: 10.3389/fpsyg.2018.02277
17. Rachels JR, Rockinson-Szapkiw AJ. The effects of a mobile gamification app on elementary students' Spanish achievement and self-efficacy. Computer Assisted Language Learning. 2018;31(1-2):72-89. doi: 10.1080/09588221.2017.1382536
18. Gao Y, Pan L. Learning English vocabulary through playing games: the gamification design of vocabulary learning applications and learner evaluations. The Language Learning Journal. 2023;51(4):451-471. doi: 10.1080/09571736.2023.2217828
19. Tamtama GIW, Suryanto P, Suyoto S. Design of English Vocabulary Mobile Apps Using Gamification: An Indonesian Case Study for Kindergarten. Int J Eng Ped. 2020;10(1):150. doi: 10.3991/ijep.v10i1.11551
20. Ishaq K, Mat Zin NA, Rosdi F, Jehanghir M, Ishaq S, Abid A. Mobile-assisted and gamification-based language learning: a systematic literature review. PeerJ Computer Science. 2021;7:e496. doi: 10.7717/peerj-cs.496
21. Bovermann K, Bastiaens TJ. Towards a motivational design? Connecting gamification user types and online learning activities. RPTEL. 2020;15(1):1. doi: 10.1186/s41039-019-0121-4
22. Sällberg H, Wang S, Numminen E. The combinatory role of online ratings and reviews in mobile app downloads: an empirical investigation of gaming and productivity apps from their initial app store launch. J Market Anal. 2023;11(3):426-442. doi: 10.1057/s41270-022-00171-w
23. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Journal of Clinical Epidemiology. 2008;61(4):344-349. doi: 10.1016/j.jclinepi.2007.11.008
24. Apple Inc. iTunes Search API documentation and iTunes RSS feeds for the App Store [Internet]. n.d. Available from: https://developer.apple.com/library/archive/documentation/AudioVideo/Conceptual/iTuneSearchAPI/
25. Lee G, Raghu TS. Determinants of Mobile Apps' Success: Evidence from the App Store Market. Journal of Management Information Systems. 2014;31(2):133-170. doi: 10.2753/mis0742-1222310206
26. Alyami M, Giri B, Alyami H, Sundram F. Social anxiety apps: a systematic review and assessment of app descriptors across mobile store platforms. Evid Based Mental Health. 2017;20(3):65-70. doi: 10.1136/eb-2017-102664
27. Alsalman D, Bu Ali ZM, Alnosaier ZF, Alotaibi NA, Alanzi TM. Gamification for Diabetes Type 1 Management: A Review of the Features of Free Apps in Google Play and App Stores. JMDH. 2020;13:425-432. doi: 10.2147/jmdh.s249664
28. Tafesse W. The effect of app store strategy on app rating: The moderating role of hedonic and utilitarian mobile apps. International Journal of Information Management. 2021;57:102299. doi: 10.1016/j.ijinfomgt.2020.102299

29. Torres-Toukoumidis A, León DV, De-Santis A, López-López PC. Gamification in Ecology-Oriented Mobile Applications—Typologies and Purposes. Societies. 2022;12(2):42. doi: 10.3390/soc12020042
30. Fijačko N, Gosak L, Cilar L, Novšak A, Creber RM, Skok P, et al. The Effects of Gamification and Oral Self-Care on Oral Hygiene in Children: Systematic Search in App Stores and Evaluation of Apps. JMIR Mhealth Uhealth. 2020;8(7):e16365. doi: 10.2196/16365
31. Brown JM, Franco-Arellano B, Froome H, Siddiqi A, Mahmood A, Arcand J. The Content, Quality, and Behavior Change Techniques in Nutrition-Themed Mobile Apps for Children in Canada: App Review and Evaluation Study. JMIR Mhealth Uhealth. 2022;10(2):e31537. doi: 10.2196/31537
32. Schmidt-Kraepelin M, Toussaint PA, Thiebes S, Hamari J, Sunyaev A. Archetypes of Gamification: Analysis of mHealth Apps. JMIR Mhealth Uhealth. 2020;8(10):e19280. doi: 10.2196/19280
33. Grabner-Hagen MM, Kingsley T. From badges to boss challenges: Gamification through need-supporting scaffolded design to instruct and motivate elementary learners. Computers and Education Open. 2023;4:100131. doi: 10.1016/j.caeo.2023.100131
34. John D, Hussin N, Zaini MK, Ametefe DS, Aliu AA, Caliskan A. Gamification Equilibrium: The Fulcrum for Balanced Intrinsic Motivation and Extrinsic Rewards in Learning Systems. IJSG. 2023;10(3):83-116. doi: 10.17083/ijsg.v10i3.633
35. Hudson G, Negbenose E, Neary M, Jansli SM, Schueller SM, Wykes T, et al. Comparing Professional and Consumer Ratings of Mental Health Apps: Mixed Methods Study. JMIR Form Res. 2022;6(9):e39813. doi: 10.2196/39813

FUNDING

No financing

CONFLICT OF INTEREST

None

AUTHORSHIP CONTRIBUTIONS

Drafting – original draft: Lucía Gerez, Verónica Daniela Larrosa Valls.
Writing–review and editing: Lucía Gerez, Verónica Daniela Larrosa Valls.