Why This App? How User Ratings, App Store Rankings, and Other App Store Information Impact Educators' Selection of Educational Apps

Emma Liptrot

Department of Educational & Counselling Psychology

McGill University, Montreal

April 2024

A thesis submitted to McGill University

in partial fulfillment of the requirements for the degree of

Master of Arts in Educational Psychology

Table of Contents

Abstract

Research indicates that educators value certain benchmarks of educational quality when choosing educational apps from app stores (i.e., curriculum, feedback, scaffolding, learning theory, and development team). However, it is unclear how other users' ratings of the app, or the app's ranking on the app store's list of "top" educational apps, might impact educators' evaluations. The present study examines how educational benchmarks, ratings, and rankings influence educators' selection of educational apps. One-hundred and fifty elementary educators viewed 18 researcher-created educational app pages and indicated their willingness to download, pay for, and rate each app. Results from a repeated-measures MANOVA and non-parametric tests revealed that educators preferred benchmark apps to buzzword apps, with a medium effect. However, they also had a strong preference for apps with positive user ratings, with a large effect, and preferred apps with a bottom ranking, with a medium effect. Educators own reasoning for whether they would download apps was also examined using word frequency analysis, corroborating the strong impact of user ratings and revealing that other aspects of app pages (e.g., visuals) may impact their app selection as well. To improve app selection, educators should rely on their own knowledge to choose apps rather than relying on user ratings or company rankings. Companies running app stores should improve their user ratings and rankings systems to facilitate selection of apps that include evidence-backed benchmarks of app quality.

Keywords: Elementary education; Mobile learning; Improving classroom teaching; Pedagogical issues; Educational apps

Abrégé

Des recherches indiquent que les éducateurs accordent d'importance aux critères de référence éducatives lorsqu'ils choisissent les applications éducatives des boutiques en ligne (c'est-à-dire : programme d'études, rétroaction, étayage, théorie de l'apprentissage et équipe de développement). Cependant, il n'est toujours pas clair quel impact les évaluations d'autres utilisateurs ou le rang de l'application dans la liste des « meilleures » applications éducatives de l'app store pourrait avoir sur les évaluations des éducateurs. La présente étude examine comment les critères de référence éducatives, les évaluations des utilisateurs et les rangs des applications impactent les sélections des éducateurs. Cent cinquante éducateurs de l'école primaire ont vu 18 pages d'applications éducatives créés par les chercheurs. Ils ont indiqué leurs volontés de télécharger, payer, et attribuer une note par étoiles pour chaque application. Les résultats d'une analyse de variance multivariée révèlent que les éducateurs ont préféré les applications avec les critères de référence éducatives aux mots tendances avec une taille d'effet moyenne. Cependant, ils ont eu une préférence forte pour les applications avec des évaluations d'utilisateurs positifs avec une taille d'effet grand et ils ont préféré les applications des derniers rangs avec une taille d'effet moyenne. Le raisonnement des éducateurs pour savoir s'ils téléchargeraient une application ou non était examiné utilisant l'analyse de la fréquence des mots. Ces analyses ont corroboré l'impact fort des évaluations d'utilisateurs et ils ont révélé que des autres aspects des pages d'applications éducatives pourraient aussi impactent leurs sélections. Pour améliorer la sélection des applications, les éducateurs devraient utiliser leur propre connaissance pour choisir les applications plutôt que les évaluations d'autres utilisateurs ou les rangs des entreprises. Les entreprises qui gèrent les boutiques en ligne d'applications devraient améliorer les systèmes

d’évaluations utilisateurs et de rang pour faciliter la sélection des applications qui incluent des critères de référence éducatives.

Mots clés : l’éducation primaire; l’apprentissage mobile; améliorer l’enseignement en classe; les issues pédagogiques; les applications éducatives

Acknowledgements

First, I would like to express my sincere thanks to my supervisor, Dr. Adam Dubé, for his endless guidance, feedback and support that were essential in the preparation of this thesis. Through every step of the way, your knowledge, dedication, and expertise have inspired me and I am sincerely grateful for all of the opportunities you have provided for me to learn and improve.

Next, I would like to thank Dr. Nikki Lobczowski for providing advice, feedback and organizing workshops that helped me to develop foundational skills essential to my research.

I would also like to thank my fellow TLC lab members, whose discussions and support have helped me to be a better researcher. A special thanks to Heather Pearson and Armaghan Montazami, whose collaboration and feedback were integral to the development of this work, and to Nandini Bharadwaj, whose friendship and mentorship has been an essential source of support throughout my degree.

To my family, thank you for all you have done for me. Mom, Dad, Grandma, Grandpa, Poppa, and Dawn, your support throughout the years has been a critical factor in my success.

Finally, thank you to my partner, Wesley who has been there for me through it all. Your love, patience, and encouragement have been a constant source of motivation for me and I am so grateful to have you by my side.

Funding for this research was provided by Social Science and Humanities Research Council.

Contributions

Much of the text from this thesis (specifically the introduction, part I, overall limitations, and conclusions) is found in a paper that was under review at the time of initial thesis submission, entitled *Why this app? How user ratings and app store rankings impact educators' selection of educational apps*, which is co-authored by Heather A. Pearson, Armaghan Montazami, and Adam K. Dubé. The paper is now published (Liptrot et al., 2024). Emma Liptrot is the first author of the paper, and was responsible for writing its contents in its entirety. Emma Liptrot was responsible for creation of study materials and data collection for the paper, while project ideation, design, and data analysis were conducted collaboratively with Armaghan Montazami and Heather Pearson, overseen by Dr. Adam Dubé. Analysis of textual data (part II) was added in addition to the contents of the paper under review. Armaghan Montazami, Heather Pearson, Jie Gao, Tania Tan, and Dr. Adam Dubé assisted in ideation, design, and coding for part II.

Liptrot, E., Pearson, H. A., Montazami, A., & Dubé, A. K. (2024). Why this app? How user ratings and app store rankings impact educators' selection of educational apps. *Computers & Education, 218*. https://doi.org/10.1016/j.compedu.2024.105080

List of Tables

List of Figures

List of Abbreviations

AI	Artificial Intelligence
ANOVA	Analysis of Variance
CK	Content Knowledge
MANOVA	Multivariate Analysis of Variance
PK	Pedagogical Knowledge
TK	Technological Knowledge
TPACK	Technological, Pedagogical, and Content Knowledge
UK	United Kingdom
US	United States

Introduction

Since the introduction of the iPad in 2010 (Apple Inc., 2010), researchers have investigated the potential for mobile devices as learning tools (e.g., Dubé et al., 2019; Fabian et al., 2016; Kim et al., 2021). Children appear to be engaged by learning with mobile technology, prompting excitement by parents, researchers, and schools at the possibility that educational apps could improve learning outcomes (Dubé et al., 2019). App developers have capitalized on the interest in educational apps; there are now over 400,000 apps in the Apple App Store's Education category (Pocket Gamer, 2024). Apple has also taken advantage of the educational interest in their mobile technologies, introducing their own applications and features to show their commitment to education (Apple Inc., 2023a). Tablet devices have become a fixture in classroom learning; in 2015, 64% of US elementary school students reported using tablets for school, and 30% used them every day (Pearson, 2015). In 2019, tablets remained the second most common digital device used in schools, behind laptops (ASCD, 2019).

Students' engagement with educational apps may help them to develop their skills and improve their competencies (Camilleri & Camilleri, 2019). In addition, students enjoy using the technology, introducing the possibility that educational apps could improve attitudes towards learning (Dündar & Akçayır, 2014). Recent meta-analyses have identified overall positive effects of apps on literacy and math achievement across studies, supporting the use of mobile devices to improve learning (Fabian et al., 2016; Kim et al., 2021). However, introducing tablets into the classroom with little to no guidance for teachers produces discouraging results: neither Bebell and Pedulla (2015) nor Carr (2012) found that the introduction of iPads improved math achievement. Dubé and colleagues (2019) propose that the effectiveness of learning with tablet devices depends on the specific apps that are used, explaining the large variation in effectiveness

between studies (Kim et al., 2021) and some discouraging results (Bebell & Pedulla, 2015; Carr, 2012). Indeed, there is huge variation in the quality of apps that are available on the Apple App Store (Dubé et al., 2020; Hirsh-Pasek et al., 2015; Meyer et al., 2021; Vaala et al., 2015). It is essential that educators can identify high-quality apps, as the success of app implementation in classrooms ultimately depends on educators' app choices. However, a lack of educational standards and useful information in the Apple App Store presents a challenge for educators (Dubé et al., 2020; Papadakis & Kalogiannakis, 2017; Taylor, Kolak, Bent, & Monaghan, 2022; Vaala et al., 2015). Moreover, educators already experience anxiety and discomfort surrounding educational technology (Fernández-Batanero et al., 2021). Understanding how educators decide which educational apps to incorporate into their teaching will identify areas where they could benefit from support. The present study, therefore, endeavours to understand how educators choose educational apps for their classrooms.

What is a "High Quality" Educational App?

A multitude of rubrics, frameworks, and checklists have been designed to evaluate educational app quality, but few are research based (Shahjad & Mustafa, 2022). Walker (2011) proposed an app evaluation rubric that formed the basis for many subsequent evaluation tools, but it lacked theoretical basis or scientific support (Shahjad & Mustafa, 2022). Other frameworks are limited in their generalizability because they are subject specific (e.g., Rosell-Aguilar, 2017) or limited to apps for preschool children (e.g., Callaghan & Reich, 2018; Kolak et al., 2021; Papadakis et al., 2017; Taylor, Kolak, Norgate, & Monaghan, 2022). Some assessment tools (e.g., Lee & Cherner, 2015; Meyer et al., 2021) provide a comprehensive way for educators to evaluate apps, but require educators to dedicate extensive time to downloading and familiarizing

themselves with each app, making them impractical for educators making initial download decisions.

To address these issues, Dubé and colleagues (2020) identified five benchmarks based on work by Vaala and colleagues (2015) and Cayton-Hodges and colleagues (2015) that educators can use as evidence-based rules of thumb when choosing apps for various subjects. These benchmarks consider the app's curricular content, pedagogical approach (feedback, scaffolding, and learning theory), and the expertise of the app's developers.

Curriculum

First, educational apps should follow a curriculum with clear learning goals (Dubé et al., 2020). Teachers look for educational apps that link to the content they teach (Falloon, 2017). Yet, many of the apps in Apple's education category do not have clear links to curriculum (Cherner et al., 2014). Further, including learning goals facilitates learning because they guide students towards meaningful learning opportunities within the app (Falloon, 2013). In general, apps with a learning goal are higher quality (Taylor, Kolak, Norgate, & Monaghan, 2022). Yet, only about half of the "top" educational apps have a clear learning goal (Taylor, Kolak, Norgate, & Monaghan, 2022), while even fewer claim to follow a particular curriculum (Dubé et al., 2020). Thus, curriculum is a useful educational benchmark to help filter out low-quality apps.

Feedback

Feedback is widely accepted as a central component of effective teaching (Petty, 2009) and has a positive impact on student learning (Wisniewski et al., 2020). Feedback that helps learners to improve their understanding has been found to support learning from educational apps (Falloon, 2013; Moyer-Packenham et al., 2018). Yet, this type of feedback is relatively rare (Callaghan & Reich, 2018; Cayton-Hodges et al., 2015; Tärning, 2018). Educators should look

for apps that provide informative feedback that can guide students towards learning and correcting their mistakes (Dubé et al., 2020).

Scaffolding

Scaffolding refers to supports that help a learner to solve a problem that they would not otherwise be able to solve on their own (Hirsh-Pasek et al., 2015; Wood et al., 1976). Scaffolding may help learners progress towards a learning goal, preventing aimless exploration of app features that may or may not be relevant (Hirsh-Pasek et al., 2015). Scaffolding improves learning from digital games (Cai et al., 2022). Educators should look for educational apps that include scaffolding such as hints and guidance, visual demonstrations and examples, or the careful sequencing of activities (Dubé et al., 2020; Callaghan & Reich, 2018; Zydney & Warner, 2016).

Learning Theory

To be effective, educational apps should be grounded in learning theory (Dubé et al., 2019; Kebritchi & Hirumi, 2008). By investigating the pedagogical approaches of 55 educational games, Kebritchi and Hirumi (2008) identified five learning theories that developers have used to guide educational game design, including direct instruction (Joyce et al., 1992), experiential learning (Dewey, 1938), discovery learning (Bruner, 1961; Ormrod, 1995), situated cognition (Brown et al., 1989), and constructivist learning (Bruckman, 1998; Kafai & Resnick, 1996). For example, a game guided by direct instruction provided opportunities for learners to practice specific mathematics concepts, while a game guided by experiential learning used role-playing to learn about global conflicts. When developers are clear that their app was designed based on a specific learning theory, it is easier for educators to decide whether it meets their goals (Dubé et

al., 2019; Dubé et al., 2020). In addition, mentioning learning theory may help educators filter out apps that are not theory-driven at all (Dubé et al., 2019).

Development Team

Including educational experts (e.g., educators or child development specialists) in an app's development team may result in the creation of more effective educational apps (Vaala et al., 2015). Educational apps should be appropriate for children's physical and intellectual needs (Callaghan & Reich 2018; Papadakis et al., 2018). Unfortunately, many educational apps include design choices that are not appropriate for the age group they target, highlighting the need for collaboration between developers and experts (Crescenzi-Lanna & Grané-Oró, 2016). Thus, educators should look for educational apps that mention collaboration with experts in child development, education, and learning content.

How Educators Judge Educational Apps

Educators' judgements of app quality should be guided by their professional knowledge. To understand teachers' knowledge of technology use in teaching, Mishra and Koehler (2008) propose the TPACK framework, wherein teachers' knowledge is conceptualized as an interaction between three components: Technological Knowledge (TK), Pedagogical Knowledge (PK), and Content Knowledge (CK). When making choices about which technology to use, these knowledge types interact to facilitate effective teaching. In the case of educational apps, educators integrate TK, PK, and CK in the intersections between these types of knowledge as they evaluate whether an app's design promotes meaningful learning of content. Dubé and colleagues' (2020) five educational benchmarks are an example of evidence that educators could look for to facilitate these judgements when choosing an educational app from app stores: *curriculum* establishes that the content is relevant; *development team* shows that the app was

created by those with pedagogical expertise; *learning theory* allows assessment of how the app can be used to achieve pedagogical goals; and *feedback* and *scaffolding* demonstrate that the app uses effective teaching methods.

On the other hand, teachers may rely on word-of-mouth when choosing educational apps. Recommendations from other teachers, students, and online reviews all impact educators' selection of digital games (Takeuchi & Vaala, 2014). Further, issues with the app selection process in the Apple App Store, as outlined in the following sections, may present a challenge for educators who attempt to rely on their own knowledge to make app choices, potentially driving them to rely on word-of-mouth indicators like user ratings and the app's ranking on Apple's "Top Charts" list of Education apps instead.

The App Selection Process

To use an educational app in their classroom, educators must first find and download it from an app store. For example, to download an educational app for use on an iPad, educators must first find it on the Apple App Store. Users can search for apps directly, using filters to narrow their search. Alternatively, they can browse through apps in the education category. However, inclusion in the education category does not guarantee educational quality. Apple's guidelines do not restrict the education category beyond requesting that developers select the "most appropriate category" for their app (Apple Inc., 2023b). As a result, education apps vary widely in pedagogical content. Some apps can barely be classified as "educational" at all (Cherner et al., 2014), while others are designed without consideration of educational theory, or lack vital learning supports (Callaghan & Reich, 2018; Dubé et al., 2020). Educators should ideally choose apps whose design and content has been supported by empirical evidence of effectiveness. Yet, the abundance of available apps and lack of standards by Apple presents a

significant challenge. Scientific research cannot keep up with the rapidly increasing selection of apps, making the comprehensive evaluation of every app impossible.

To help parents and educators narrow their search, several websites and blogs review apps. While these websites are popular tools for educators and parents, they lack scientific backing (Hirsh-Pasek et al., 2015). Some websites include sponsored content, introducing bias that educators may struggle to detect (List et al., 2022). Even unsponsored app review websites like Common Sense Media may not recommend apps with high educational potential (Taylor, Kolak, Bent, & Monaghan, 2022). Consequently, educators cannot rely solely on app review websites. They must use the information provided by the App Store to judge whether an app meets their needs and goals. Every app available in the Apple App Store has an app page, which contains information to help consumers make download decisions (see Fig. 1). The following sections will outline the various information available on these app pages and describe what is known regarding how this information may impact educators' download decisions.

Visual Information

Visual information for each app consists of an icon and a selection of images. Visual design may impact educators' download decisions; several studies have found that app icons influence download decisions (Cao et al., 2021; 2022; Jylhä & Hamari, 2019; Wang & Li, 2017) and consumers may judge the quality of an app based on elements of its icon (Jylhä & Hamari, 2019). In general, consumers prefer designs that they consider aesthetically pleasing, but preferences for specific aspects of design are inconsistent and vary for different age groups and demographics (Wang & Lin, 2019). For example, Wang and Lin (2019) found that older children (ages 9-11) prefer website designs with a medium level of *visual complexity* (defined by the level of visual detail and number of elements on the page), whereas younger children (ages 7-8) prefer

a low level. Yet, in the App Store, apps designed for the youngest category of children (under 5) are the most visually complex (Dubé et al., 2020).

Despite consumers' preference for certain aspects of visual design, there is no evidence that more visually appealing apps or games result in improved learning outcomes. Javora and colleagues (2018) tested 9-11 year old children's learning from a high aesthetic and low aesthetic version of a science learning game. While children preferred the high aesthetic version, opting to use it when given the choice, children who played the high aesthetic version of the game did not demonstrate better learning outcomes. The authors concluded that aesthetics may be less important in a formal learning context (i.e., a classroom setting) where children do not have the option to choose a more attractive design. Consequently, visual design may not be an important consideration for educators. Indeed, Montazami and colleagues (2022a) tracked educators' eye movements as they viewed various educational app pages and found that educators spent more time evaluating the text descriptions of the apps of an app than the images, which may suggest that textual information is more carefully considered by educators.

Textual Information: Educational Buzzwords vs. Benchmarks

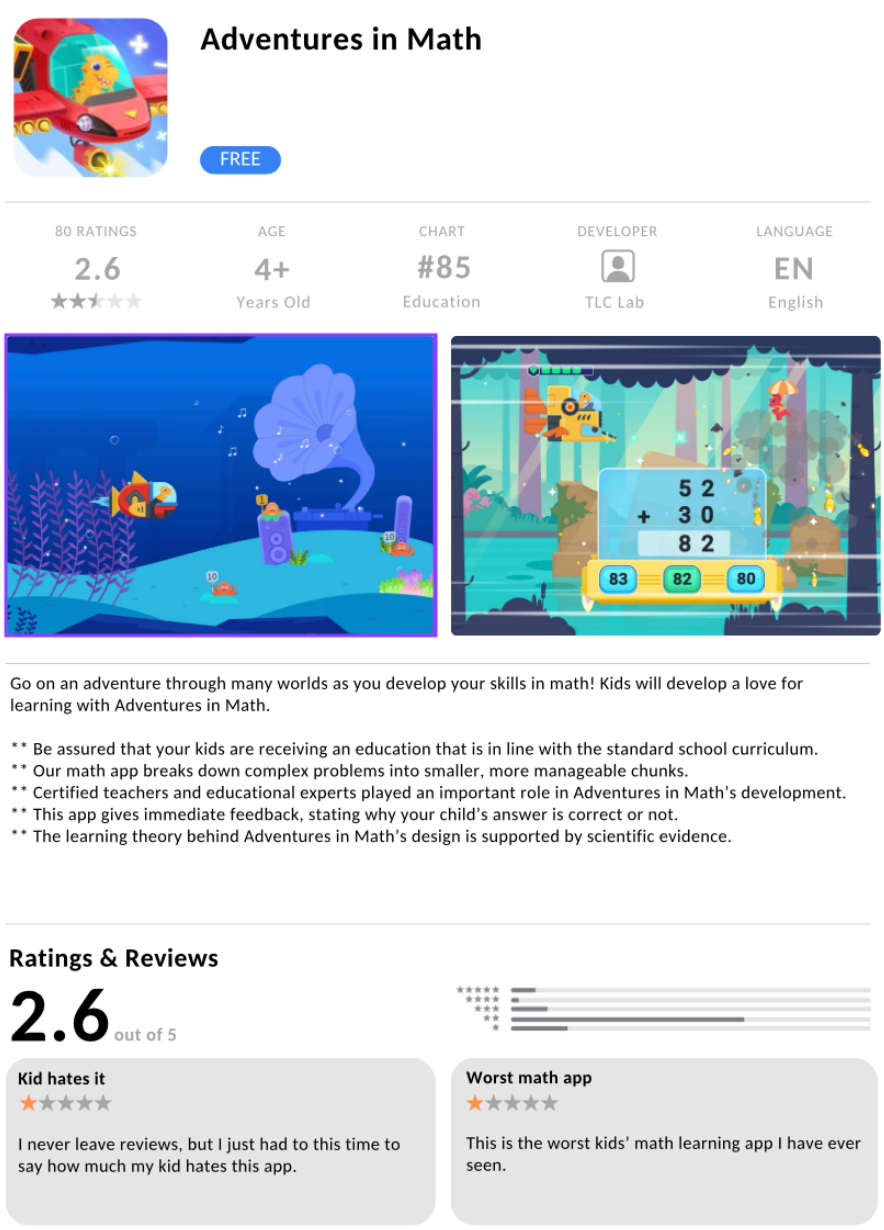
Each app page includes a written description where developers can describe their app. Written descriptions should ideally provide information about the design of the app, curriculum, pedagogical approach, and learning supports. However, the detail included in written descriptions varies widely, with reports of descriptions ranging from 13 to 1089 words in length (Vaala et al., 2015). In addition, many developers treat app descriptions as an opportunity to promote their product (Larkin, 2013), enticing consumers with catchy language rather than providing useful information about the app's educational quality. Websites and blogs encourage developers to attract customers with apps that are engaging, interactive, personalized,

multimedia, and hands-on (e.g., Guru Technolabs, 2023; Kim, 2023; Manchanda, 2022). These "buzzwords" are also common on popular education blogs (e.g., Farber, 2015; Common Sense Education, n.d.). However, they do not provide enough information to help educators distinguish high-quality apps from low-quality apps. Apps, by their very nature, include multimedia, are personalized, interactive, and hands-on, and any app could describe itself as engaging. Thus, educators should not rely on buzzwords when choosing educational apps.

In contrast to buzzwords, educational benchmarks (curriculum, feedback, scaffolding, learning theory, development team) provide evidence of educational quality when they are included in descriptions (Dubé et al., 2020). These benchmarks provide information about the app's content and pedagogy, allowing educators to use their Technological, Pedagogical, and Content Knowledge to judge app quality. As such, if educators are using their professional knowledge to make app decisions, they may prefer apps that include educational benchmarks. To examine whether educators prefer benchmarks, Montazami and colleagues (2022a) created a series of fake app pages whose written descriptions mentioned either educational benchmarks or buzzwords (i.e., engaging, interactive, hands-on, personalized, and multimedia) and asked educators to evaluate each app. Educators were more likely to download apps with educational benchmarks, gave them a higher rating, and were willing to pay more for them compared to buzzword apps. These results demonstrate that educators value educational benchmarks when making app decisions, which may suggest that they are relying on their professional knowledge to choose effective apps. Unfortunately, few real apps mention benchmarks in their text descriptions (Cayton-Hodges et al., 2015; Dubé et al., 2020; Vaala et al., 2015), which may limit educators' ability to use their professional knowledge, prompting them to look for other signals of app quality.

Figure 1

Example of an App Page as they Appear in the Apple App Store



Note. This figure shows a fake app page that was created for the purposes of this study but was formatted to replicate a real app page in the App Store.

Ratings and Reviews

The App Store allows users to rate their experience with an app on a scale from 1 (most negative) to 5 (most positive). Users can also provide a short review in the form of written feedback if desired. An average star rating is prominently displayed on search pages, making them a particularly salient cue for potential users. Previous studies of apps in various categories (Biviji et al., 2020; Burgers et al., 2016; Krishnan & Selvam, 2019) have shown that consumers prefer apps that have been rated positively and rated many times, perhaps because they view ratings as indicators of social desirability and quality (Burgers et al., 2016). Although these studies did not examine educational apps specifically, educators may also rely on user ratings to signal the quality of an educational app.

Unfortunately, user ratings may not be useful for establishing educational quality. Dubé and colleagues (2020) examined 90 of the top education apps in the Apple App Store and concluded that ratings do not provide useful information for parents or educators. In their study, an app's rating was not related to the number of educational benchmarks in the app's description. Ratings may be useful for evaluating educational quality in combination with reviews if the reviews mention educational benchmarks or specific features of the app. However, most app reviews only provide general praise (Pagano & Maalej, 2013). Singh and Suri (2022) identified themes in the user reviews of four educational apps on the Google Play store and explored how these themes relate to star ratings. They found that mentioning content quality and teaching quality in app reviews was significantly, but weakly, associated with a user's rating of an app. However, several other themes were also related to star ratings, including technical quality and customer support quality, which may make it difficult for educators to determine why an app has received a particular rating. Furthermore, most reviews were positive, and the authors noted that

"learners rate the apps high, even when they perceive the technical quality and content quality to be medium" (Singh & Suri, 2022, p. 7), which may make it difficult for educators to use app ratings to distinguish the highest quality apps. Even when app reviews do provide information related to an app's educational content, anyone can review an app, while educators have years of training and experience that they should use to make their own judgements. Thus, educators may use ratings and reviews as word-of-mouth cues to the desirability of an app but should avoid relying on these signals to assess educational quality.

Rankings

When consumers choose to browse a specific category in the App Store, Apple provides a "Top Charts" list of apps in that category. An app's rank denotes its position on the "Top Charts" list for its category. Rankings impact the order in which apps are shown when consumers browse through a category, such that top-ranking apps (closer to 1) are shown first. Rank is also displayed on app pages, referred to as the "chart number" (see Fig. 1). App market researchers have determined that ratings, number of downloads, and app usage are all factors that impact chart position, but Apple has not disclosed specifics of how their algorithm determines app ranking (Blacker, 2022; Walz, 2015). In any case, Apple's rankings are a measure of an app's popularity, which does not necessarily align with educational quality. Still, educators may view popularity as a proxy for quality, resulting in a preference for top-ranked apps. Indeed, consumers are willing to pay more for apps that are ranked closer to the top of bestseller lists (Carare, 2012).

Researchers across various fields have called the quality of apps in Apple's lists of top-ranked apps into question, pointing out that many make misleading claims and lack scientific basis (Tongdee & Markowitz, 2018; Wisniewski et al., 2019). Education apps are no different:

few of the top mathematics (Dubé et al., 2020) and literacy (Vaala et al., 2015) apps mention following a specific curriculum, and an app's rank is not associated with the number of educational benchmarks in its description (Dubé et al., 2020). Furthermore, only some app pages present a ranking at all. These issues, along with the lack of transparency surrounding how rank is determined, make ranking an unreliable way to assess app quality. Thus, like user ratings, educators should avoid relying on rankings when making educational app decisions.

The Current Study

Montazami and colleagues (2022a) determined that educators look for five educational benchmarks when choosing apps (curriculum, feedback, scaffolding, learning theory, development team), suggesting that they rely on their professional knowledge to choose high-quality educational apps. However, other information could impact educators' selection as well. Relying on information that is not reliable for determining app quality (e.g., ratings and rankings; Dubé et al., 2020) could result in the selection of poor quality educational apps. Consequently, students' learning could suffer. Given the prevalence of tablet learning in today's schools (Pearson, 2015), it is essential that educators choose high-quality educational apps. Thus, the current study builds on previous work by considering how information besides educational benchmarks impacts educators' evaluations of an app. The study is divided into two parts. Part I aims to understand how educational benchmarks, user ratings, and Apple rankings impact educators' selection of educational apps using a repeated-measures design. In part I, educators evaluate 18 researcher-fabricated educational apps, manipulated to vary by rating, ranking, and whether they contain educational benchmarks. Part I addresses a limitation of previous work by investigating how others' judgements (i.e., user ratings and Apple rankings) impact educators' app decisions. Part II aims to understand educators' own reasons for downloading educational

apps by using a text mining approach to analyze their text responses when asked to explain their reasoning for why they would or would not download each of the apps from part I. Part II will help to contextualize educators' app preferences and identify other potential influences on educators app selection that warrant future study.

Part I

Part I investigates educators' selection of educational apps by varying the ratings and rankings of 18 researcher-fabricated educational apps, in addition to whether they contain educational benchmarks or buzzwords. Part I is guided by the following research questions:

RQ1: Do educators prefer apps with educational benchmarks over educational buzzwords? Based on previous findings by Montazami and colleagues (2022a), it is predicted that educators will prefer (i.e., be more likely to download, pay more for, and rate higher) apps whose descriptions mention educational benchmarks (cf., buzzwords).

RQ2: Does an app's user rating impact educators' evaluation of the app? It is predicted that educators will prefer apps with positive user ratings, based on evidence that consumers in general prefer highly rated apps (Biviji et al., 2020; Burgers et al., 2016; Krishnan & Selvam, 2019).

RQ3: Does an app's Apple ranking impact educators' evaluation of the app? Based on previous evidence that consumers prefer to download apps near the top of bestseller charts (Carare, 2012), it is predicted that educators will prefer top-ranked apps.

RQ4: Does the presence of educational benchmarks interact with rating and ranking to influence app evaluations? In response to RQ1, RQ2, and RQ3, it is expected that educational benchmarks, rating, and ranking will all factor into educators' download decisions. Yet, no prior research to our knowledge has considered how this information may be used in combination to

influence consumers' decisions. Thus, RQ4 explores how educational benchmarks, ratings, and rankings interact to impact educators' app evaluations. Due to the lack of preceding studies, RQ4 has no specific hypotheses.

Method

Participants

A power analysis was conducted using More Power (Campbell & Thompson, 2012) and G*Power with conservative effect sizes at 80% power. Neither software had the capability to perform power analyses with the specific analysis used in this study (three-way repeated measures MANOVA), but by comparing results for related tests, a sample size of 150 was determined to be sufficient. Participants were recruited through Prolific, an online participant pool that allows researchers to target a specific sample using filters. Participants were only included in the study if they self-reported that they were currently teaching students between grades one and six in one of three English-speaking countries with similar learning contexts: Canada, the United States, or the United Kingdom. Compared to other online participant pools, Prolific produces high quality data (Peer et al., 2021). Still, participants were required to pass three out of four attention checks embedded within the study to ensure quality of response. Three participants did not pass the attention checks and were excluded from the final sample.

The final sample included 150 educators (33% male) from Canada (10%), the US (38%), and the UK (52%) with a mean age of 39.71 years (SD = 10.17) and an average 12.33 years of teaching experience (SD = 8.33). Most participants were white (87%) and taught at a public school (88%). The grades participants taught were diverse, with grade 6 being the most common (39%). Most participants reported using apps in their classrooms (94%), most commonly using tablets (63%) and laptops (63%). Frequency of educational app use varied, with a few times a

week being the most common response (43%). See Table 1 for more information about participant demographics. All participants consented to participation in the study, which was approved by the university's research ethics board.

Table 1

Participant Demographics

	n	%
Gender		
Woman	101	67
Man	49	33
Ethnicity		
White/European	130	87
Southeast Asian	10	7
Black/African/Caribbean	7	5
South Asian	5	3
Latin American	4	3
Arab	1	1
Indigenous	0	0
Other	1	1
Country		
United Kingdom	78	52
United States	57	38
Canada	15	10
School/home Area		
Suburban	67	45
Urban	43	29
Town or village	32	21
Rural	8	5
School type		
Public	132	88
Private	14	9
Other	4	3
Reported use of apps for learning in the classroom		
Yes	141	94
No	9	6
Frequency of educational app use		
Every day	36	24
A few times/week	65	43
About once/week	20	13
A few times a month	10	7
Once/month	5	3
Less than once/month	7	5
Never/Not applicable	7	5

Technology used for educational apps		
Tablets	95	63
Laptops	95	63
Desktops	39	26
Cell phones	16	11
Other	5	3
None/not applicable	6	4
Grades taught by participant*		
Kindergarten	46	31
Grade 1	25	17
Grade 2	40	27
Grade 3	38	25
Grade 4	39	26
Grade 5	57	38
Grade 6	59	39

Note: Some questions allowed participants to choose multiple responses. Percentages may not always add up to 100%.
*In the UK, children start first grade earlier than in the US and Canada. UK participants' responses were converted to correspond with the North American system.

Procedure

The study was administered online via Qualtrics. Participants provided consent, then answered demographic questions including details about their teaching role and their use of technology in the classroom. Next, participants viewed and evaluated 18 researcher-generated images designed to imitate app pages in the Apple App Store. A 2x3x3 repeated measures design was deployed with the 18 apps varying along three factors: app type (benchmark or buzzword), rating (negative, neutral, or positive), and ranking (bottom, middle, or top). The app pages were presented individually, in a random order, asking participants to evaluate them as if they were choosing an app to use for kindergarten to grade 3 students.

Materials

The stimuli for this study consisted of 18 researcher-created app pages designed to replicate the experience of looking in the Apple App Store on an iPad to choose an educational app. Each page was randomly assigned an icon and two images from actual educational apps, with a fictitious title and one-sentence generic description of the app to match. Figure 2 shows an

example of an app page that was created for this study and illustrates how the page was manipulated to vary by app type (educational benchmarks or buzzwords), user rating (negative, neutral, and positive), and ranking (bottom, middle, or top). All other elements of the app pages were randomized or held constant to reduce potential confounds. *T*-tests and one-way ANOVAs were conducted to confirm that the image complexity (as defined and calculated by Dubé et al., 2019) did not differ by app type, $t(16) = 0.259$, $p = .799$, rating, $F(2, 15) = 0.90$, $p = .429$, or ranking $F(2, 15) = 0.208$, $p = .815$. All apps were listed as "free" and the age, developer, and language information were the same across apps. See Appendix A for the full list of apps.

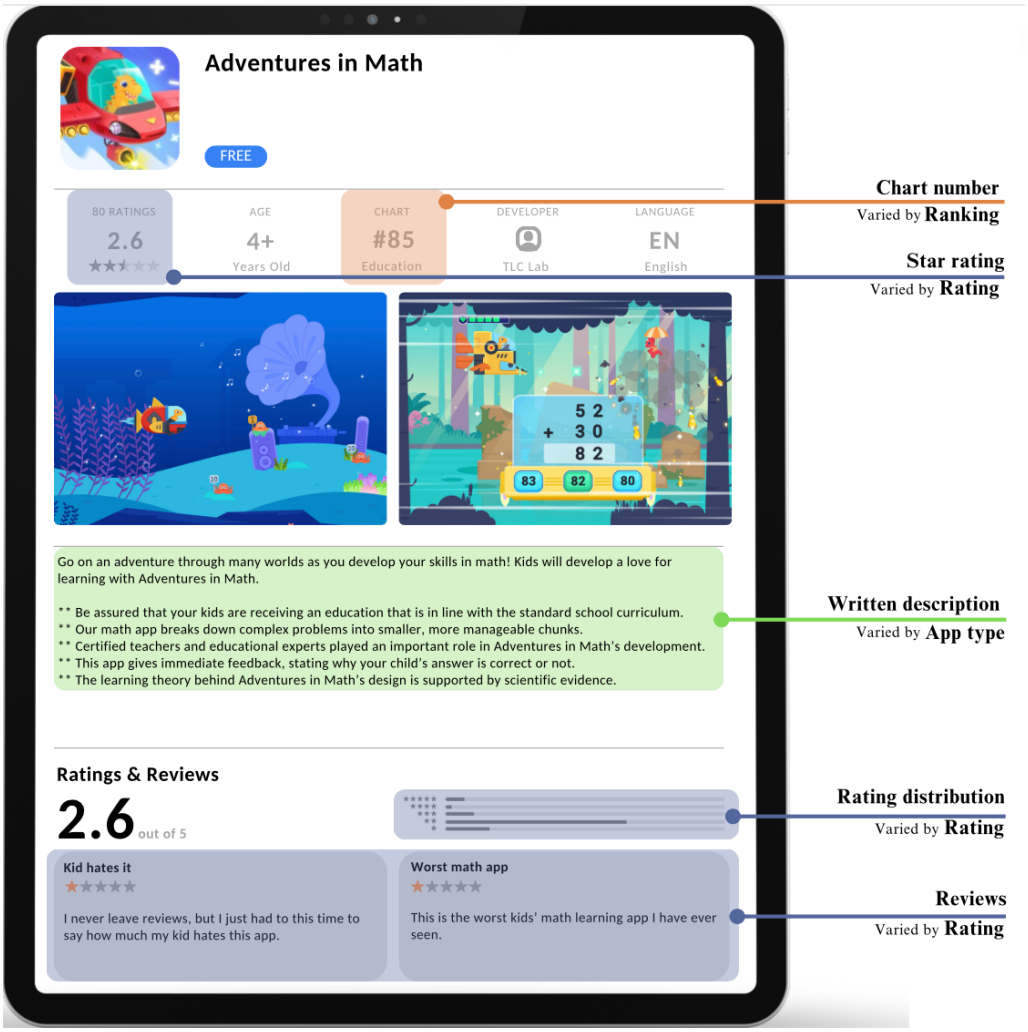
App Type

The written description on each app page was manipulated to contain either educational benchmarks or buzzwords. For benchmark apps, the app description included five one-sentence bullet points which mentioned each of Dubé and colleagues' (2020) educational benchmarks: curriculum, scaffolding, feedback, development team, and learning theory. For buzzword apps, the bullet points mentioned: engaging, interactive, hands-on, personalized, and multimedia.

To make descriptions different enough so that participants could not easily detect a pattern and identify the app condition, an AI language model (chatGPT; OpenAI, 2022) was used to help generate a pool of sentences for each benchmark and buzzword that varied in phrasing, but met criteria set by the researcher. These sentences were then curated and randomly assigned such that each app page contained one sentence for each benchmark/buzzword. The order in which the benchmarks/buzzwords were presented was randomized for each app. Description lengths (number of words) did not differ by rating, $F(2, 15) = 1.49$, $p = .257$, ranking $F(2, 15) = 0.187$, $p = .831$, or app type $t(16) = -1.98$, $p = .065$.

Figure 2

Example of a Fake Educational App Page.



Note: Coloured sections show where the app page was manipulated to vary by app type, rating, and ranking. The app page depicted in this figure is benchmark app with a negative rating and bottom ranking.

Rating

In the App Store, each app page displays an average rating out of 5, along with a distribution of all past user ratings and a selection of written reviews (see Figure 2). In the present study, the numerical and written ratings were manipulated such that each app could be categorized as negative, neutral, or positive. McIlroy and colleagues (2016) classified ratings below three stars as negative, three stars as neutral, and four- and five-star ratings as positive. Following these groupings while ensuring clear distinctions between categories, negative, neutral, and positive apps were assigned ratings from 1.0 to 2.9, 3.4 to 3.9, and 4.4 to 4.9 stars, respectively. This information was displayed at the top of the page and in the "ratings & reviews" section at the bottom of the page. Further, the visual distribution of ratings was manipulated to skew towards each app's assigned star rating. Rating did not differ by app type, $t(16) = -.160$, $p = .875$ or ranking $F(2, 15) = .056$, $p = .956$.

The written reviews were manipulated such that negative apps included two one-star reviews with negative commentary, neutral apps included a three-star and a four-star review with neutral commentary, and positive apps included two five-star reviews with positive commentary. Each app included one generic review that rehashed the user's rating in word form (e.g., "this is the best math learning app ever!") and one review that discussed the user's experience with the app (e.g., "My kids are always asking to play this app."), representing two of the most common types of app reviews (Maalej et al., 2016). These reviews included no details about the app itself (i.e., no mention of features or bugs). The length of reviews did not differ by app type, $t(16) = 0.091$, $p = .929$, rating, $F(2, 15) = 0.91$, $p = .424$, or ranking $F(2, 15) = 0.03$, $p = .973$.

Ranking

An app's "chart number" represents its ranking on Apple's list of "Top Charts" within the education category. Thus, rankings closer to one are "higher." For the present study, apps with a chart number between 1-10 were considered top-ranking apps, apps with a chart number between 10-50 were middle-ranking, and apps with a chart number between 50-90 were bottom-ranking. To ensure clear distinctions between categories, top-ranking apps were assigned a random ranking from 3 to 9, middle-ranking apps were assigned a ranking from 43 to 49, and bottom-ranking apps were assigned a random ranking from 83 to 89. Rankings did not differ by app type, $t(16) = -.122$, $p = .904$ or rating $F(2, 15) = .003$, $p = .997$.

Measures

Participants answered four questions for each app, three of which have been used in prior studies (Montazami et al., 2022a; 2022b; Pearson et al., 2022). These questions reflect decisions educators would make if they were choosing apps from the App Store.

Would you download this app? (Yes/No). Previous studies have assessed educators' app evaluations by asking them whether they would download an app (Montazami et al., 2022a; 2022b; Pearson et al., 2022). Choosing to download an app is the first step towards its use in the classroom. Whether an educator would download an app should indicate whether the app meets their minimum standards for educational use. Thus, this measure was treated as an overall evaluation, while other measures were used to gain a more detailed understanding of educators' preferences.

How likely would you be to download this app? (1, extremely unlikely, to 7, extremely likely). Like the previous measure, this question is designed to understand whether an app meets educators' standards for classroom use. This measure was not included in previous studies

(Montazami et al., 2022a; 2022b; Pearson et al., 2022), but was added to the present study to provide a continuous measure of educators' download decisions.

Approximately how much would you be willing to pay for this app? ($0 to $30)

Although all the apps in this study were free, most of the "top" educational apps in the Apple App store are not (Dubé et al., 2020). The average price of a top educational app is $14.48 CAD, suggesting that consumers are willing to pay for educational apps (Dubé et al., 2020). According to Furner and Zinko (2018), consumers undergo cost/benefit analysis when deciding how much to pay for an app, and how much participants are willing to pay directly relates to their evaluation of what an app is "worth." Thus, how much educators are willing to pay for an app was used as an app evaluation measure, as in previous studies (Montazami et al., 2022a; 2022b; Pearson et al., 2022). Participants could enter any whole number from $0 to $30.

Overall, how would you rate this app? (1 to 5 stars) As previously discussed, app ratings reflect users' general positive or negative experiences with an app. As in previous studies (Montazami et al., 2022a; 2022b; Pearson et al., 2022), participants were asked to rate each app based on its app page.

Statistical Analysis

Data were screened for deviations from normality and outliers. How much educators were willing to pay (cost) was found to be skewed and kurtotic for several conditions. Thus, a logarithmic transformation with a base of 10 was applied to cost prior to MANOVA analysis. Following transformation of cost, there were still some deviations from normality exceeding Hair and colleagues' (2010) recommended skewness cut-off. However, deviations were less extreme, and MANOVA is robust to moderate deviations from normality (Olson, 1976).

Examination of Z-scores and Mahalanobis distances revealed univariate (n = 28) and multivariate (n = 4) outliers. Inconsistencies in participant data indicated that four of these values were clearly the result of an error in data collection and were thus treated as missing data. The missing values were replaced with the mean score for the measure (Tabachnick & Fidell, 2007). The remaining outliers were not removed from the analysis, as they may represent meaningful natural variation; several researchers recommend against removing outliers without reason because doing so may inflate Type I error rates (Bakker & Wicherts, 2014; Karch, 2023; Wilcox, 1998).

The impact of app type, rating, and ranking on whether educators would download an app (Yes/No) was examined using non-parametric tests that are robust to outliers (Scheff, 2016). The impact of app type on whether educators would download an educational app was examined with a related samples Wilcoxon signed rank test. The impact of rating and ranking on whether educators would download an app was examined with Friedman's related-samples analysis of variance by ranks, followed by Wilcoxon signed rank tests.

The impact of app type, rating, and ranking on educators' preferences was further examined in a three-way repeated-measures MANOVA with educators' willingness to download an app (download likelihood), pay for the app (cost), and rate the app (rate) as dependent variables. Removing all outliers was not found to impact overall results. Thus, 24 outliers were retained in the MANOVA analysis.

Results

RQ1: Do Educators Prefer Apps with Educational Benchmarks Over Educational Buzzwords?

A Wilcoxon signed-rank test of, "would you download this app? Yes/No" revealed that educators would choose to download benchmark apps (M = 57.11%, SD = 18.43) significantly more often than buzzword apps (M = 49.11%, SD = 18.31), $Z = 4.89$, $p < .001$, $r = 0.41$.

The MANOVA also found significant differences in educators' preferences depending on app type, Wilk's $\Lambda = 0.88$, $F(3,147) = 6.93$, $p < .001$, with a medium effect size, $\eta^2_p = .12$. Bonferroni-corrected ANOVAs were conducted as follow-up for each of the three dependent measures in the analysis: download likelihood, cost, and rate (see Table 2). There were medium effects of app type on download likelihood, cost, and rating (all $p < .001$), with educators consistently preferring benchmark apps over buzzword apps. See Table 3 for means.

RQ2: Does an App's User Rating Impact Educators' Evaluation of the App?

Friedman's related-samples analysis of variance by ranks revealed that app rating significantly impacted whether educators would download an app (Yes/No), $\chi^2(2) = 251.23$, $p < .001$, with a strong level of agreement, Kendall's $W = .84$. Follow-up Wilcoxon signed-rank tests revealed that educators would download apps with positive ratings more frequently than negative ratings ($Z = 10.62$, $p < .001$, $r = .87$) or neutral ratings ($Z = 9.09$, $p < .001$, $r = .74$), and would download apps with neutral ratings more frequently than negative ratings ($Z = 9.85$, $p < .001$, $r = .80$). See Table 3 for means.

The MANOVA found that rating had a main effect on educators' app evaluations, Wilk's $\Lambda = 0.15$, $F(6, 592) = 154.39$, $p < .001$, with a large effect size, $\eta^2_p = .61$. Follow-up repeated measures ANOVAs revealed that rating had a large effect on download likelihood, cost, and

educators' own ratings (all p <.001, see table 2). Pairwise comparisons revealed that educators were more likely to download, pay more for, and rate apps higher with positive ratings than neutral (all p <.001) or negative (all p <.001) ratings at a Bonferroni corrected alpha level of .006 (.017/3). See Table 3 for means.

Table 2

Within-Subjects Effects from Follow-Up ANOVA Analyses

ANOVA	Effect	*df*	*F*	*p*	η^2_p
Download Likelihood	App type	1, 149	11.45	<.001*	.07
	Rating**	2, 267	641.29	<.001*	.81
	Ranking	2, 298	14.23	<.001*	.09
	App type by rating	2, 298	56.57	<.001*	.28
	App type by ranking	2, 298	24.77	<.001*	.14
Cost	App type	1, 149	11.60	<.001*	.07
	Rating**	2, 232	209.78	<.001*	.59
	Ranking	2, 298	5.48	.005*	.04
	App type by rating**	2, 282	17.14	<.001*	.10
	App type by ranking	2, 298	0.99	.005*	.01
Rate	App type	1, 149	20.86	<.001*	.12
	Rating**	2, 226	726.67	<.001*	.83
	Ranking	2, 298	14.15	<.001*	.09
	App type by rating	2, 298	19.28	<.001*	.12
	App type by ranking	2, 298	11.90	<.001*	.07

Note: Significance evaluated at the Bonferroni-adjusted alpha level p < .017 (.05/3).
**Assessed using Greenhouse-Geisser statistic because the assumption of sphericity was violated.

Table 3

Means and Standard Deviations

Measure	Condition	*M*	*SD*	*SE*
Download	App type			
	Buzzword	49.11%	18.31	1.49
	Benchmark	57.11%	18.43	1.51
	Rating			
	Negative	12.78%	20.01	1.63
	Neutral	56.78%	29.92	2.44
	Positive	89.78%	18.09	1.48
	Ranking			
	Bottom	54.33%	20.20	1.65
	Middle	53.67%	19.39	1.58
	Top	51.33%	18.37	1.50
Download Likelihood	App type			
	Buzzword	3.77	0.77	.06
	Benchmark	3.96	0.79	.06
	Rating			
	Negative	1.98	0.90	.07
	Neutral	4.07	1.06	.09
	Positive	5.54	1.02	.08
	Ranking			
	Bottom	4.01	0.85	.07
	Middle	3.87	0.81	.07
	Top	3.71	0.76	.06
Cost	App type			
	Buzzword	2.42	3.21	.26
	Benchmark	2.72	3.78	.31
	Rating			
	Negative	0.58	1.45	.12
	Neutral	2.07	3.45	.28
	Positive	5.05	6.31	.52
	Ranking			
	Bottom	2.72	3.75	.31
	Middle	2.43	3.26	.27
	Top	2.55	3.48	.29
Rate	App type			
	Buzzword	2.77	0.44	.04
	Benchmark	2.90	0.45	.04

Rating			
Negative	1.67	0.54	.04
Neutral	2.91	0.51	.04
Positive	3.91	0.68	.06
Ranking			
Bottom	2.93	0.51	.04
Middle	2.78	0.45	.04
Top	2.78	0.44	.04

RQ3: Does an App's Apple Ranking Impact Educators' Evaluation of the App?

Friedman's related-samples analysis of variance by ranks revealed no significant difference in whether educators would download an app by ranking, $\chi^2(2) = 4.43$, $p = .109$, Kendall's $W = .02$.

Ranking had an overall main effect on educators' app evaluations, Wilk's $\Lambda = 0.86$, $F(6, 592) = 7.86$, $p < .001$, with a medium effect size, $\eta^2_p = .07$. Follow-up Bonferroni-corrected ANOVAs revealed that ranking had a medium effect on downloads ($p < .001$, see Table 2). Pairwise comparisons revealed that educators were less likely to download top-ranking apps than bottom-ranking apps ($p < .001$) or middle-ranking apps ($p = .003$), but there was no difference between middle and bottom-ranking apps ($p = .020$) at the Bonferroni corrected alpha level of .006 (.017/3). See Table 3 for means. Similarly, ranking had a small effect on cost ($p = .005$), whereby educators were willing to pay more for bottom-ranking apps than middle-ranking apps ($p = .001$), but there were no significant differences between bottom and top-ranking apps ($p = .015$) or middle and top-ranking apps ($p = .680$) at the Bonferroni-corrected alpha level of .006 (.017/3). Educators' own ratings of the app were significantly affected by the app's ranking with a medium effect size ($p < .001$), such that educators rated bottom-ranking apps higher than middle or top-ranking apps (both $p < .001$), but there was no difference between middle and top-ranking apps ($p = .893$).

RQ4: Does the Presence of Educational Benchmarks Interact with Rating and Ranking to Influence App Evaluations?

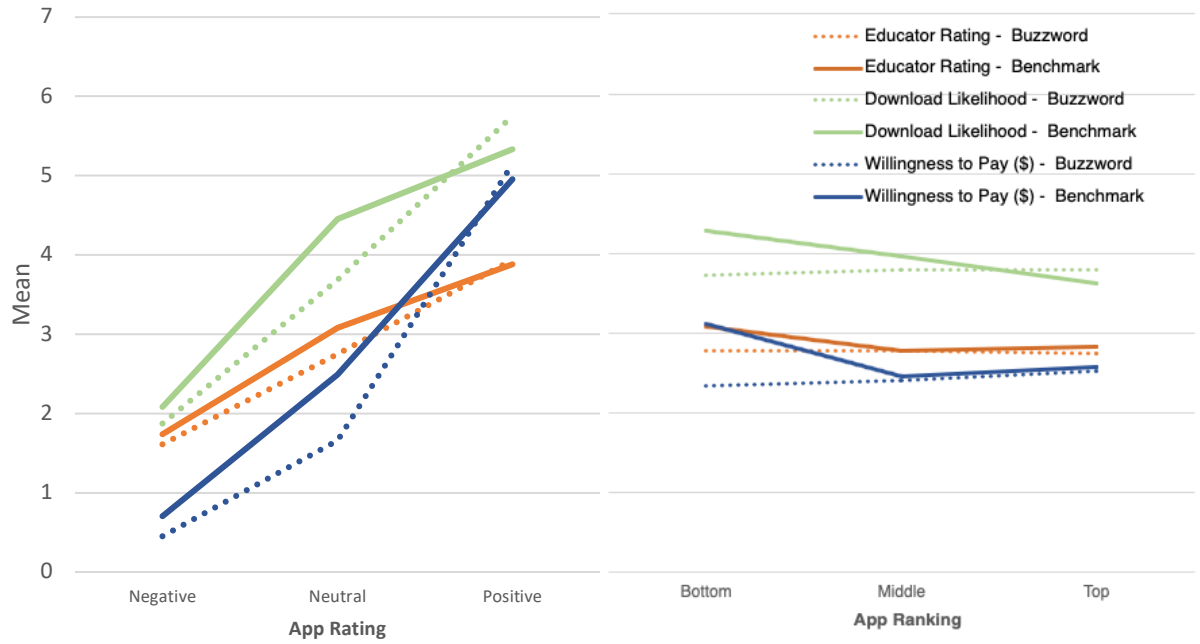
The MANOVA revealed a significant app type by rating interaction, Wilk's $\Lambda = 0.72$, $F(6, 592) = 17.79$, $p < .001$, with a large effect size, $\eta^2_p = .15$ (see Figure 3). Follow-up ANOVAs revealed significant app type by rating interactions for download likelihood, cost, and rate (all $p < .001$, see Table 2). Examination of the simple main effects for each dependent variable revealed that whether educators preferred benchmark or buzzword apps depended on the app's rating. Educators were only willing to pay more for benchmark apps and rate benchmark apps higher when the app's rating was neutral (both $p < .001$), while there was no difference between benchmark and buzzword apps for apps with a negative or positive rating with a Bonferroni-adjusted alpha level of .002 (.017/6). For apps with a neutral rating, educators preferred to download benchmark apps, but for apps with a positive rating, they were more likely to download buzzword apps (both $p < .001$).

The app type by ranking interaction was also significant, Wilk's $\Lambda = 0.79$, $F(6, 592) = 12.29$, $p < .001$, with a medium effect size, $\eta^2_p = .11$ (see Figure 3). Follow-up ANOVAs revealed this interaction was significant for download likelihood ($p < .001$), cost ($p = .005$), and rating ($p < .001$), see Table 2. For all three dependent measures, educators only preferred benchmark apps when apps had a bottom ranking (all $p < .001$). Differences between buzzword and benchmark apps were not significant for middle or top-ranking apps at a Bonferroni-corrected alpha level of .006 (.017/6). Further, whether ranking impacted educators' preferences depended on the app type. For benchmark apps, educators were more likely to download bottom-ranking apps than middle or top-ranking apps, and more likely to download middle than top-ranking apps (all $p < .001$); they also were willing to pay more for bottom-ranking apps than middle or top-ranking

apps, and rated bottom-ranking apps more highly (all p <.001). Yet, for buzzword apps, ranking made no difference.

Figure 3

Educator App Evaluations: App Type by Rating (Left), App Type by Ranking (Right)



Note: Solid lines represent Benchmark apps while dotted lines represent Buzzword apps. Although displayed here on the same scale, download likelihood could range from 1-7, educator rating from 1-5, and willingness to pay from $0-30. Figure depicts overall trends across all variables with significant differences noted in the text. Error bars not included due to overlap among error bars from different dependent variables.

Part I Discussion

Overall, part I suggests that educational benchmarks, app ratings, and Apple rankings all impact educators' evaluations of educational apps.

Educational Benchmarks

Supporting our hypotheses, educators were more likely to download apps if their descriptions included benchmarks compared to buzzwords. They were also willing to pay more for apps with benchmarks and give benchmark apps a higher rating, supporting previous findings by Montazami and colleagues (2022a) that educators prefer apps that mention educational

buzzwords. Mishra and Koehler (2008) propose that educators use TPACK to judge how technology interacts with pedagogy to convey content effectively. Benchmarks provide information that educators need to facilitate these judgements, in contrast to buzzwords, which provide little useful information for educators. Thus, this result may suggest that educators use their professional knowledge to make decisions about educational apps.

Ratings

Although educators had an overall preference for apps including benchmarks, the present study also revealed that educational benchmarks interacted with user ratings, such that educators only preferred apps with educational benchmarks for apps with a neutral rating. One possible explanation is that educators rely primarily on user ratings but look for benchmarks when ratings are ambiguous. Supporting this notion, the effects of rating were consistently very large compared to the effects of app type. For example, educators said they would download apps with educational benchmarks 57% of the time, compared to buzzwords 49% of the time, while they would download apps with positive ratings 90% of the time, compared to negative ratings 13% of the time. In general, educators had a strong preference for apps with positive ratings: they were more likely to download them, pay more for them, and rate them positively.

Educators' preference for apps with positive ratings supports our hypotheses, and corroborates previous work (Biviji et al., 2020; Burgers et al., 2016; Krishnan & Selvam, 2019) that established consumers' preference for apps with positive ratings. It is possible that educators interpreted negative ratings as indicators of problems with the app (e.g., advertisements, in-app purchases, privacy concerns, or bugs) that were not considered in this study. However, educators should be more discriminating with positively-rated apps. Educators' reliance on ratings may indicate that they follow the judgements of others rather than using their own professional

knowledge. All the reviews included on the app pages were phrased in such a way that they were either ambiguous or appeared to be written by a parent; meaning, educators had no reason to believe reviewers had educational expertise. Furthermore, the contents of the reviews did not include specific information about the app's educational contents. Thus, educators would be better served to use their own expertise to select apps, as Dubé and colleagues (2020) posit that ratings are not a useful way to evaluate educational quality.

Rankings

Ranking had a significant impact on educators' decisions to download, pay for, and rate an app. However, contrary to our expectations, educators generally preferred apps ranked toward the bottom of the education charts (i.e., apps at chart position 83-89 rather than apps in the "top 10"). Dubé and colleagues (2020) pointed out that rankings may be confusing, as they are presented without context. It is possible that educators misunderstood ranking, interpreting higher numbers as a "good" ranking, when in fact "top" apps have rankings closer to one, which could explain the contradictory results. Alternatively, educators may not consider rankings at all, and differences could instead be explained by other features of the app pages such as visual characteristics. Although the visual complexity of images did not differ across conditions, educators may have had preferences for visual characteristics that were not controlled for in the present study. In contrast to Carare and colleagues (2012), who determined that consumers would pay more for top-ranked apps using app store data, our study showed educators individual app pages. Ranking may be more important when educators are searching for apps, as it impacts the order in which apps are shown, making top-ranked apps more salient. Investigation of educators' own reasons for whether they would download an app may provide further insight

into the extent to which they consider rankings in their evaluations and reveal other considerations that impact their decisions as well.

Limitations

An important limitation to note in part I is the presence of some outliers and deviations from normality in the data. While we did not find it appropriate to remove outliers from our analyses, MANOVA assumes there are no outliers (Meyers et al., 2016) suggesting the need for caution. However, non-parametric test results corroborated the MANOVA results, enhancing credibility. Future research that replicates aspects of our methodology could further substantiate our findings. More stringent directions for participants, such as providing a smaller range for how much they are willing to pay, may prevent some outliers.

Part II

Part I revealed that benchmarks, ratings, and rankings may all contribute to educators' decisions to download an educational app. However, it remains unclear how other considerations, including those that were not controlled by the researcher in part I, may impact educators' decisions. Consequently, an additional aim emerged throughout the research process; part II aims to identify common topics in educators' own reasoning for why they would or would not download each of the apps in part I. Part II is guided by the following exploratory research question:

RQ5: What are the most common reason educators give for whether they would download an educational app? Based on the results of part I, it is predicted that benchmarks, ratings, and rankings will be common reasons. Additional reasons could include visual characteristics (Cao et al., 2021; 2022; Jylhä & Hamari, 2019; Wang & Li, 2017). RQ5 will

explore which reasons are most common a) overall; b) for benchmark apps; c) for buzzword apps; d) for apps they would download; and e) for apps they would not download.

Method

Data

Part II uses data from the same participants and procedure at part I, including 2700 open-ended responses to the question: *please describe your reasoning for whether you would download this app or not*. In addition to the app evaluation questions described in part I, after viewing each app, participants were asked to describe their reasoning for whether they would download the app in an open-ended response. To promote more detailed responses, participants were required to write at least 50 characters (approximately 1-2 sentences).

Identification of Reasons

Word frequency analysis was used to identify themes within text data from 2700 responses to the question: *please describe your reasoning for whether you would download this app or not*, addressing RQ5. Frequency analysis is a form of content analysis that counts the occurrence of topics within a set of text data (Mayring, 2015). The text data was cleaned, prepared, and a list of the most frequent words within the dataset was generated. Next, researchers categorized the most frequent terms through an iterative coding process.

Data Cleaning and Preparation. First, Microsoft Word's editing feature was used to identify and correct spelling mistakes within the data. All mistakes were reviewed and only corrected if the intended spelling was obvious. Grammar mistakes were not corrected, with the exception of word choice errors (e.g., *doe snot* was corrected to *does not*). All words were corrected to American standard spelling to account for regional variations in spelling. Next, data was imported to R Statistical Software (v4.3.2; R Core Team, 2023) and processed using code

adapted from the LASER Institute's text mining workshop (LASER Institute, 2021). Packages used to process the text included tidytext (Silge & Robinson, 2016) and tidyverse (Wickham et al., 2019). The text was converted to lowercase and split into individual words, then all numbers and white spaces were removed. A list of stop words (i.e., words that are not useful such as "it" or "and") was removed from the data using the tidytext package (Silge & Robinson, 2016). Prior to removal, two researchers reviewed the list and identified words that may be meaningful in the context of the present study; these words were retained in the analyses. Educators used a total of 17905 words in their responses, excluding stop words. However, some words were repeated many times, while others were only used once. Thus, a list of "unique" terms and their associated frequencies was generated. The final word list included 1998 unique words, meaning participants used a total of 1998 different terms, excluding stop words.

Word Classification. All words with a frequency greater than 10 were sorted into categories. A minimum frequency of 10 was established to ensure that all the included words represented a pattern in the data, and to ensure there were enough instances of each word's use to facilitate interpretation of the word in context. For example, a word with a frequency less than 10 was mentioned in less than .4% of the 2700 responses.

A total of 285 words had a frequency of 10 or greater, comprising 14% of the 1998 "unique" words from the text. However, the 285 words that were categorized accounted for the majority of the total text in educators' responses because they were the most frequently used terms. For example, the word "reviews" alone accounted for 7% of the total text because it was used 1310 times, out of 17905 words in total. When the frequencies of the words that were categorized are taken into account, 78% of the text was categorized.

Table 4

Category Definitions and Examples

Category	Definition	Number of Words*	Sample Words**	Example in Context
Reviews/ratings	Words that indicate the participant is referring to the ratings or reviews of the app. In some cases, this could be referring to the reviewer (e.g., parents).	22	reviews, ratings, low, high, star	"Going by the reviews, I'm not going to install it."
Ranking	Words that refer to an app's chart position/ranking.	1	chart	"It didn't chart very well. As a result, I wouldn't really go for it."
Theme	Words that refer to the thematic elements of the app, including setting, story, or characters.	9	characters, theme, zombie, space, dogs	"The characters on the app are also super cute and the pupils would love them."
Visuals	Words that refer to the visual characteristics of the app, including the images, interface, or visual design.	19	graphics, images, interface, colorful, design	"The app doesn't look exciting enough and seems quite boring with basic graphics."
Written information	Words that indicate the user is looking at the written information (description) on the app page.	3	description, blurb, write	"The description is informative and sounds reliable."
Benchmark	Words that indicate the user is considering an	23	feedback, curriculum,	"It's been developed with child and

	educational benchmark: curriculum, feedback, scaffolding, learning theory, development team. Should reflect the language we used to communicate benchmarks in the app pages or a synonym. Can list members of a development team.		developed, research, aligned	educational experts and provided targeted feedback which would show the children where they went wrong if necessary.”
Buzzword	A direct variation of one of the five educational buzzwords used in the study.	7	engaging, interactive, personalized, hands, multimedia	“It looks very engaging and interactive, whilst also consolidating the students' mathematical knowledge.”
Academic application/outcome	Words that refer to a possible academic application of the app or anticipated outcome. For example, may refer to the academic content of the app, mention of the expected user (e.g., "students") and their anticipated reaction to it.	45	children, students, math, kids, learning	“I couldn't tell how they were supposed to be helping aid in the kid's math understanding.”
Suitability	Words that indicate the user is judging the suitability of an app for use by a particular group of students, such as consideration of	14	age, younger, suitable, appropriate, level	“It looks great for younger children and would be suited to this age group.”

	age-appropriateness and difficulty.			
Judgement	Words that indicate a positive, negative, or neutral judgement about the app's desirability or qualities. The judgement describes an evaluation of the app itself (typically an adjective) rather than an anticipated outcome.	52	good, great, fun, bad, poor	“The app doesn't look bad or good, it just doesn't appeal to me and I think I could find something better.”
Cost/benefit	Words that indicate user is weighing the costs and benefits of the app, determining its worth.	8	time, free, worth, waste, pay	“I doubt I would download it as other people mention there are better apps so it would be a waste of time and money.”
More information	Words indicate that they are requesting more information to make a decision.	7	try, information, trial, explore, check	“I would try a demo or a trial, seems worth taking a look.”
Gamification	Words that indicate that the user views the app as gamified or game-like.	4	games, play	“There's a variety of games that kids can use to learn math.”

Note:
*Total number of unique words coded to each category
**Sample of the 5 words with the highest frequency count in each category, excluding variations of the same word.

To create the categories, four researchers first reviewed approximately 2.5% of the words from the text, beginning with the most frequent words, and independently formulated categories based on common topics. Next, the four researchers agreed upon a list of categories and definitions through discussion. Two coders categorized approximately the next 2.5% of most frequent words using the established categories and definitions, then resolved any discrepancies and revised the categories through discussion. This process was repeated until both coders agreed on over 80% of word classifications, at which point the primary coder completed the categorization of all words with a frequency of over 10. Any previously categorized words were re-coded by the primary coder using the final coding scheme to account for any changes that may have occurred throughout the category formation process. As a final measure of reliability, a second coder categorized a random sample of 10% of the words, revealing substantial inter-rater reliability, as demonstrated by Cohen's kappa coefficient of .80 (Landis & Koch, 1977).

A total of 13 categories were defined; reviews/ratings, ranking, theme, visuals, written information, benchmarks, buzzwords, academic applications/outcomes, suitability, judgement, cost/benefit, more information, and gamification (see table 4 for definitions and examples). Some categories were guided by the research questions of part I (i.e., benchmark, buzzword, rating, and ranking categories). Others emerged through identification of common topics in the text. Coders assigned each word to a single category. If the appropriate category was not immediately obvious, coders reviewed samples of the word in context to inform their decision. In some cases, a word could fall into any of several categories depending on context. To resolve such issues, whichever context was the most common was used to determine the word's category. In cases where there was no clear majority usage, a word was deemed "uninformative." "Uninformative" words were defined as "words that are not informative, usually because they

restate words from the question or are used in too many different contexts to distinguish meaning" (e.g., "app," "download," or "content"). Uninformative words were not considered in analyses. A total of 71 words were deemed uninformative, resulting in a total of 214 categorized words included in the final analysis. See Appendix B for the full list of words and their associated categories.

Measures

Reasons

To determine the most common reasons educators give for whether they would download an educational app, the percentage of educators' overall responses that mentioned each of the 13 categories defined in table 4 was calculated (RQ5a). Whether each response included at least one word in each category was first determined using R, then exported to SPSS Statistics using the haven package (Wickham et al., 2023), where the percentages of responses including a word in each category were calculated.

Educators' reasons were then compiled separately for buzzword apps (RQ5b) and benchmark apps (RQ5c). To accomplish this, the number of responses for each app type that included a specific reason was divided by the total number of responses for that app type. For example, if an educator mentioned a word in the "theme" category for 3 out of the 9 benchmark apps, this would indicate that they cited "theme" as a reason for whether they would download a benchmark app 33% of the time.

Finally, educators' reasons were compiled separately for apps educators would download (RQ5d), and apps educators would not download (RQ5e). Note that the total number of apps educators would download was different for each participant, depending on their responses to the question, *would you download this app?* as described in part I. On average, participants indicated

that they would download 10 out of 18 apps, but participants could download as many as 18 or as few as zero.

Statistical Analysis

To determine the most common reasons educators give for whether they would download an educational app, Friedman's 2-way repeated-measures ANOVAs were conducted due to severe deviations from normality in the data. To reduce the overall number of comparisons, stepwise step-down analysis was used for post-hoc analyses. Stepwise step-down comparisons were conducted using SPSS Statistical Software using an extension of the procedure described by Campbell and Skillings (1985), as described in the IBM SPSS Statistics Algorithms manual (IBM, 2022). Rather than conducting all possible pairwise comparisons, the stepwise step-down analyses order the reasons by their rank (as determined by the Friedman's test), then identify groups of reasons that are not significantly different from each other (i.e., homogeneous subgroups).

Results

RQ5a: What Are the Overall Most Common Reasons Educators Give for Whether they would Download an Educational App?

A Friedman's 2-way repeated measures ANOVA revealed statistically significant differences in how frequently the various reasons were mentioned, $\chi^2(12) = 1024.058$, $p < .001$. Stepwise step-down post-hoc analyses found that judgement formed the highest-ranked homogeneous subgroup on its own, indicating that it was significantly more common than all other reasons. Reviews/ratings and academic applications/outcomes formed the next-highest ranked homogeneous subgroup ($p = .567$). These reasons were significantly less common than judgement, but more common than all other subgroups. Visuals was the next highest-ranked

reason, followed by a subgroup including benchmarks, buzzwords, theme, cost/benefit, more information, and suitability (p = .186). Next, gamification and written information formed a subgroup (p = .691). Finally, ranking formed the lowest-ranked subgroup on its own. Results of the post-hoc analyses are shown in table 5. Means and standard deviations are displayed in figure 4.

Table 5

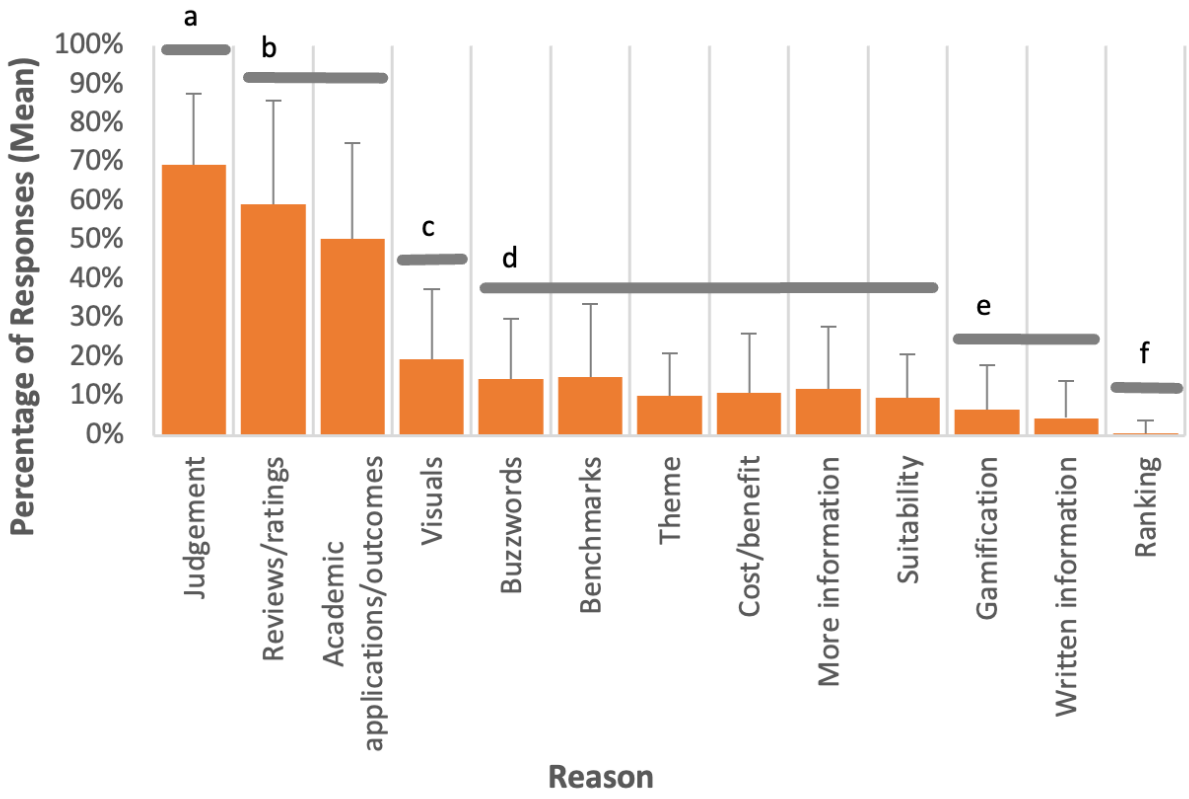
Homogeneous Subgroups of Overall Reasons

	Subgroup					
	f	**e**	**d**	**c**	**b**	**a**
Ranking	3.07					
Written Information		4.167				
Gamification		4.827				
Suitability			5.783			
More information			5.943			
Cost/benefit			5.947			
Theme			6.03			
Benchmarks			6.5			
Buzzwords			6.56			
Visuals				7.597		
Academic applications/outcomes					11.07	
Reviews/ratings					11.31	
Judgement						12.197
Test Statistic		1.93	9.51		2.41	
Adjusted Sig. (2-sided)		0.691	0.186		0.567	

Note: Table cells include rank scores as determined by Friedman's related-samples analysis of variance by ranks. Homogenous subgroups are groups of reasons that do not differ significantly.

Figure 4

Overall Most Common Reasons



Note: Error bars display one standard deviation above the mean. Lines above each bar indicate homogeneous subgroups. All reasons under the same bar are not statistically different from one another as determined by a Friedman's test using stepwise step-down comparisons. Reasons are organized in descending order by rank score as determined by the Friedman's test. Letters are assigned in order of the highest-ranked subgroup (a) to the lowest-ranked subgroup (f).

RQ5b: What Are the Most Common Reasons for Whether Educators Would Download a Benchmark App?

The Friedman's test revealed statistically significant differences among reasons for benchmark apps. $\chi^2(12) = 973.43$, $p < .001$. Stepwise step-down post-hoc analyses found that there was some overlap among subgroups, meaning some reasons belonged to multiple subgroups. Results of the post-hoc analyses are shown in Table 6. Means and standard deviations

are displayed in figure 5. Much like educators' overall reasons, judgement was the most common reason for benchmark apps and was significantly more common than all other subgroups. Some other notable results include that, unlike in educators' overall responses, buzzwords and benchmarks belonged to distinct homogeneous subgroups, with benchmarks belonging to a higher-ranked homogeneous subgroup. This indicates that for benchmark apps, benchmarks were a more common reason for downloading an app than buzzwords.

RQ5c: What Are the Most Common Reasons for Whether Educators Would Download a Buzzword App?

The Friedman's test revealed statistically significant differences among reasons for buzzword apps. $\chi^2(12) = 1013.80$, $p < .001$. Much like educators' reasons for benchmark apps, stepwise step-down post-hoc analyses revealed some overlap among subgroups. Results of the post-hoc analyses are shown in Table 7. Means and standard deviations are displayed in figure 6. Notably, buzzwords and benchmarks belonged to distinct homogeneous subgroups, with buzzwords belonging to higher-ranked homogeneous subgroups than benchmarks. This indicates that for buzzword apps, buzzwords were a more common reason for downloading an app than benchmarks.

Table 6

Homogeneous Subgroups of Reasons for Benchmark Apps

	Subgroup							
	h	**g**	**f**	**e**	**d**	**c**	**b**	**a**
Ranking	3.56							
Written information	4.59	4.59						
Gamification		4.90	4.90					
Suitability			5.68	5.68				
Buzzwords			5.71	5.71				
More information			5.90	5.90				
Cost/benefit				5.94	5.94			
Theme				6.36	6.36	6.36		
Visuals					7.01	7.01		
Benchmark						7.28		
Reviews/ratings							11.07	
Academic applications/ outcomes							11.15	
Judgement								11.86
Test Statistic	6.41	0.67	9.57	5.35	7.11	4.66	0.17	
Adjusted Sig. (2-sided)	0.072	0.969	0.072	0.532	0.118	0.358	0.999	

Note: Table cells include rank scores as determined by Friedman's related-samples analysis of variance by ranks. Homogenous subgroups are groups of reasons that do not differ significantly.

Table 7

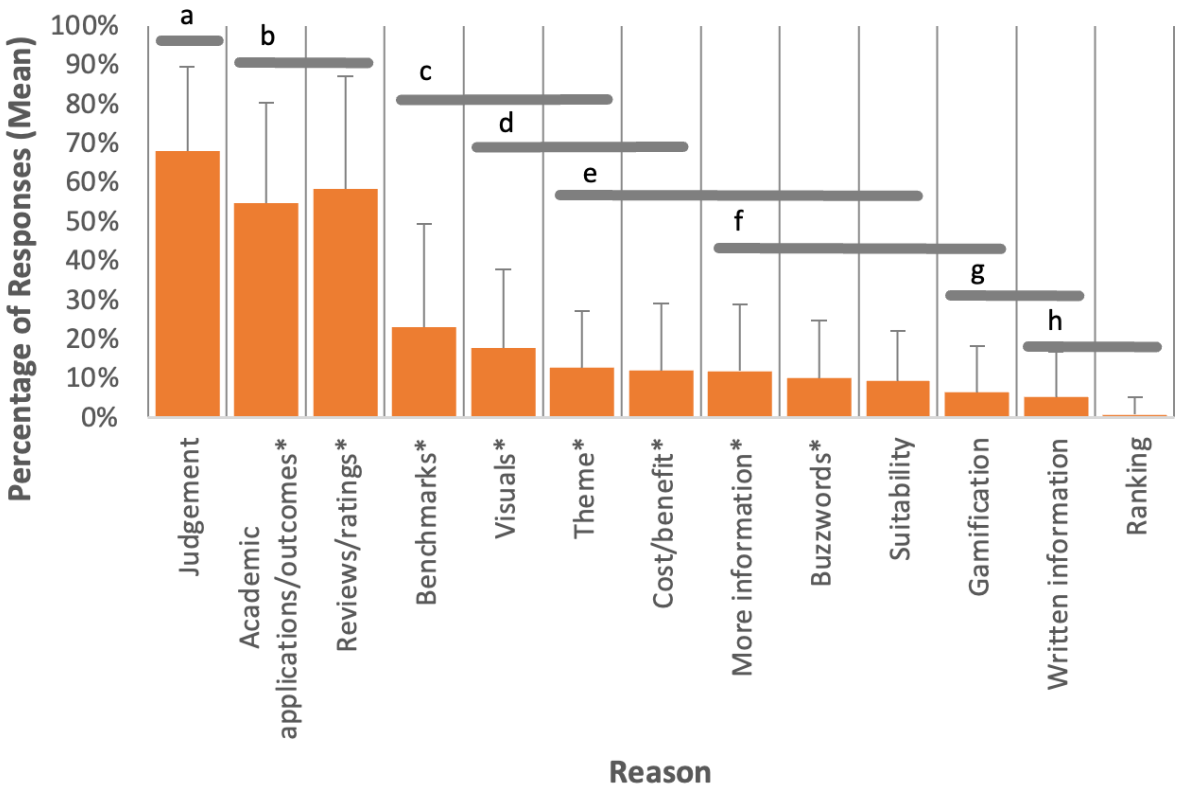
Homogeneous Subgroups of Reasons for Buzzword Apps

	Subset								
	i	**h**	**g**	**f**	**e**	**d**	**c**	**b**	**a**
Ranking	3.75								
Written information	4.44	4.44							
Benchmarks		4.93	4.93						
Gamification		5.28	5.28	5.28					
Theme			5.66	5.66					
Cost/benefit			5.77	5.77					
Suitability			5.87	5.87					
More information				6.11	6.11				
Buzzwords					7.23	7.23			
Visuals						7.90			
Academic applications /outcomes							10.59		
Reviews/ratings								11.28	
Judgement									12.19
Test Statistic	3.53	4.44	9.82	5.51	5.61	0.67			
Adjusted Sig. (2-sided)	0.333	0.392	0.109	0.508	0.111	0.969			

Note: Table cells include rank scores as determined by Friedman's related-samples analysis of variance by ranks. Homogenous subgroups are groups of reasons that do not differ significantly.

Figure 5

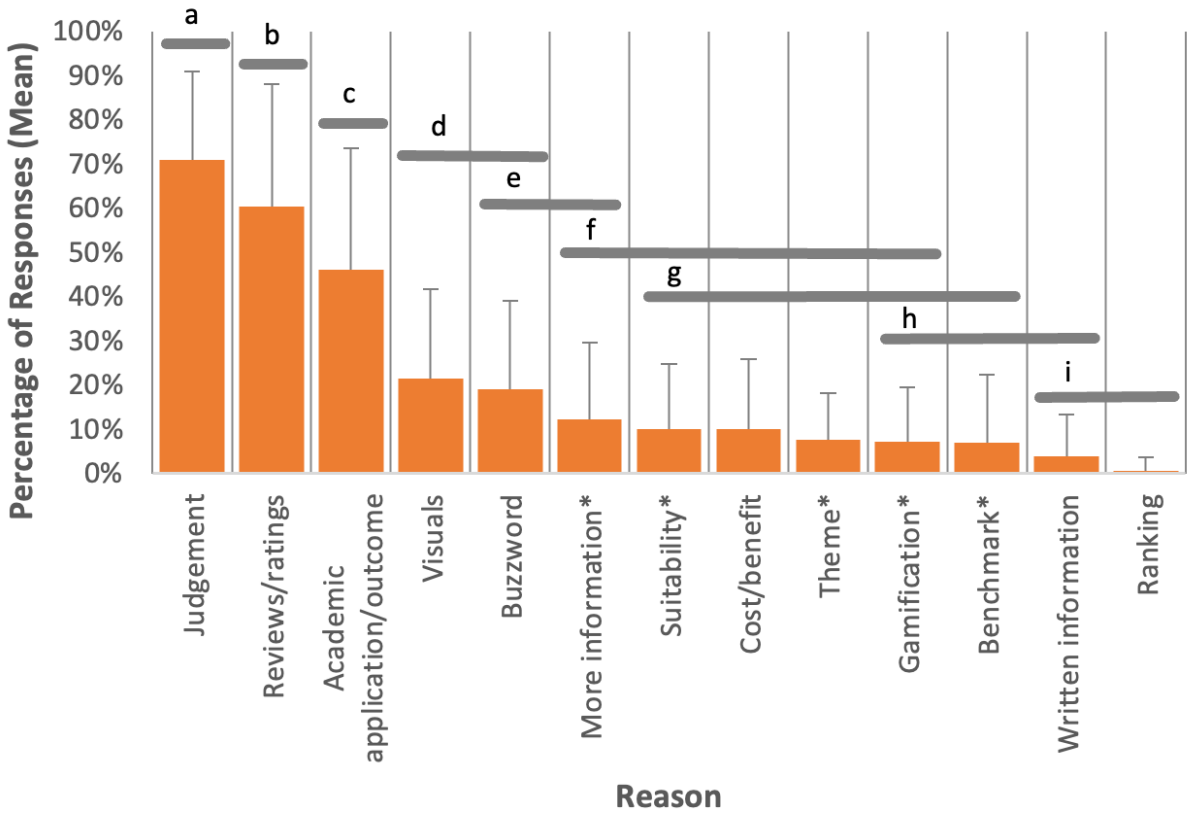
Most Common Reasons for Benchmark Apps



Note: Error bars display one standard deviation above the mean. Lines above each bar indicate homogeneous subgroups. All reasons under the same bar are not statistically different from one another as determined by a Friedman's test using stepwise step-down comparisons. Reasons are organized in descending order by rank score as determined by the Friedman's test. Letters are assigned in order of the highest-ranked subgroup (a) to the lowest-ranked subgroup (h). *Starred reasons indicate that the reason's rank order as determined by the Friedman's test is different than the order for educators' overall reasons. Note that direct comparisons were not performed.

Figure 6

Most Common Reasons for Buzzword Apps



Note: Error bars display one standard deviation above the mean. Lines above each bar indicate homogeneous subgroups. All reasons under the same bar are not statistically different from one another as determined by a Friedman's test using stepwise step-down comparisons. Reasons are organized in descending order by rank score as determined by the Friedman's test. Letters are assigned in order of the highest-ranked subgroup (a) to the lowest-ranked subgroup (i). *Starred reasons indicate that the reason's rank order as determined by the Friedman's test is different than the order for educators' overall reasons. Note that direct comparisons were not performed.

RQ5d: What Are the Most Common Reasons Educators Provide for Apps they Would Download

Among only the apps educators indicated they would download, a Friedman's test revealed significant differences among reasons, $\chi^2(12) = 944.75$, $p < .001$. Results of stepwise

step-down pairwise comparisons are shown in table 8. Means and standard deviations are displayed in figure 7. There was some overlap among subgroups, meaning some reasons belonged to multiple subgroups. For example, judgement and academic application/outcomes formed the highest ranked homogeneous subgroup (p = .133), followed by a subgroup consisting of academic applications/outcomes and reviews/ratings (p = .949). Notably, unlike in educators' overall responses, visuals formed a subgroup with benchmarks, buzzwords, and more information (p = .078), indicating no significant differences in how frequently these reasons were mentioned.

Table 8

Homogeneous Subgroups of Reasons for Apps Educators Would Download

	Subgroup							
	h	**g**	**f**	**e**	**d**	**c**	**b**	**a**
Ranking	3.40							
Written information	4.24	4.24						
Gamification		5.01	5.01					
Cost/benefit			5.43	5.43				
Suitability			5.63	5.63				
Theme			5.98	5.98	5.98			
More information				6.30	6.30	6.30		
Buzzwords					6.94	6.94		
Benchmark					7.01	7.01		
Visuals						7.35		
Reviews/ratings							10.59	
Academic applications/outcomes							11.24	11.24
Judgement								11.87
Test Statistic	5.64	3.25	8.49	6.28	8.95	9.37	0.81	5.26
Adjusted Sig. (2-sided)	0.108	0.383	0.115	0.287	0.094	0.078	0.949	0.133

Note: Table cells include rank scores as determined by Friedman's related-samples analysis of variance by ranks. Homogenous subgroups are groups of reasons that do not differ significantly.

RQ5e: What Are the Most Common Reasons Educators Provide for Apps they Would Not Download?

The Friedman's test revealed statistically significant differences in how frequently educators mentioned the various reasons for apps they would not download $\chi^2(12) = 984.87$, $p < .001$. Results of stepwise step-down pairwise comparisons are shown in table 9. Means and standard deviations are displayed in figure 8. Judgement and reviews/ratings formed the highest-ranked homogeneous subgroup ($p = .998$) and were significantly more common than all other reasons.

Table 9

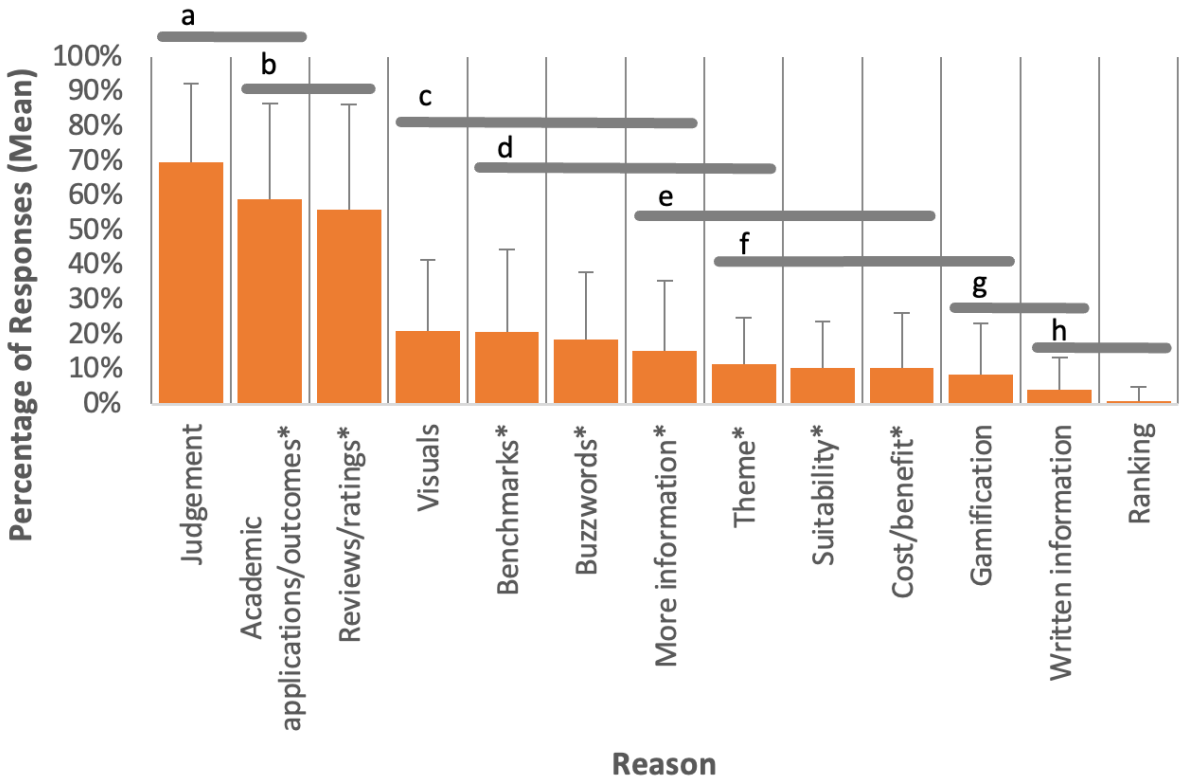
Homogeneous Subgroups of Reasons for Apps Educators Would Not Download

	Subgroup					
	f	**e**	**d**	**c**	**b**	**a**
Ranking	4.14					
Written information	4.94	4.94				
Gamification	5.14	5.14	5.14			
More information		5.51	5.51			
Benchmark		5.60	5.60			
Buzzwords		5.91	5.91			
Suitability		6.02	6.02			
Cost/benefit			6.05			
Theme			6.10			
Visuals				7.76		
Academic applications /outcomes					10.15	
Reviews/ratings						11.64
Judgement						12.04
Test Statistic	7.62	12.37	10.27			0.24
Adjusted Sig. (2-sided)	0.092	0.064	0.201			0.998

Note: Table cells include rank scores as determined by Friedman's related-samples analysis of variance by ranks. Homogenous subgroups are groups of reasons that do not differ significantly.

Figure 7

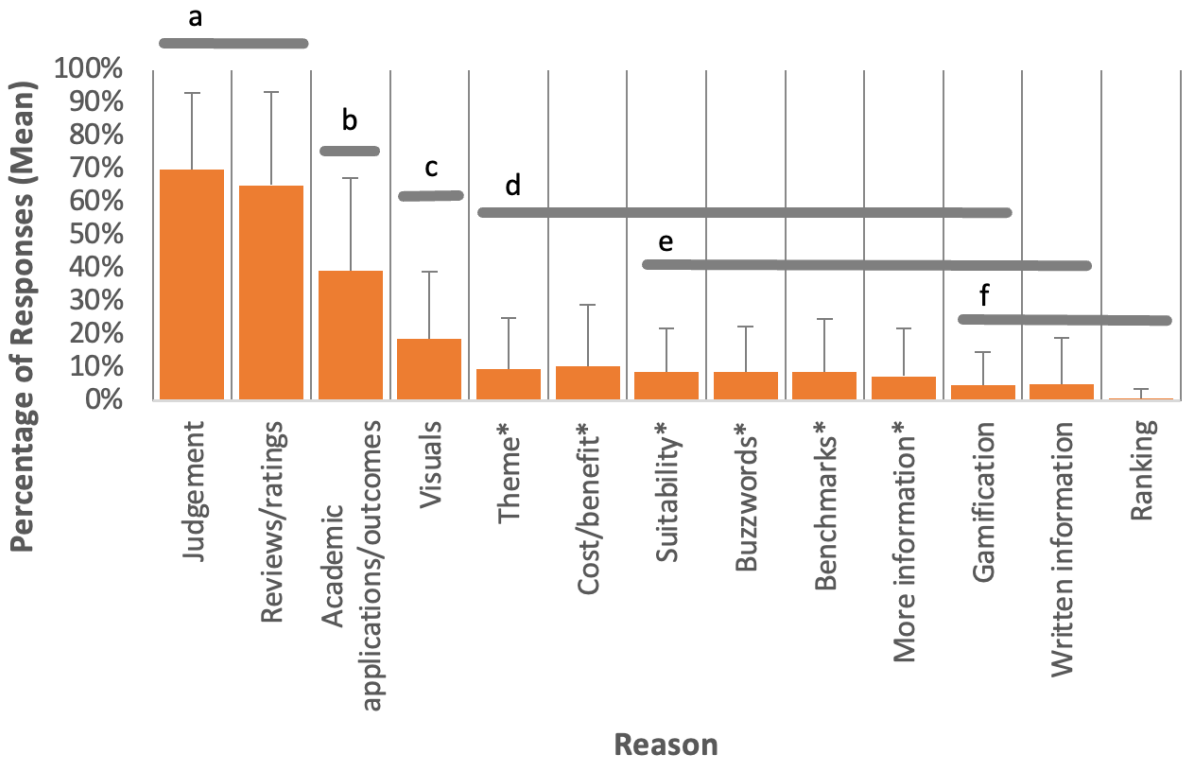
Most Common Reasons for Apps Educators Would Download



Note: Error bars display one standard deviation above the mean. Lines above each bar indicate homogeneous subgroups. All reasons under the same bar are not statistically different from one another as determined by a Friedman's test using stepwise step-down comparisons. Reasons are organized in descending order by rank score as determined by the Friedman's test. Letters are assigned in order of the highest-ranked subgroup (a) to the lowest-ranked subgroup (h). *Starred reasons indicate that the reason's rank order as determined by the Friedman's test is different than the order for educators' overall reasons. Note that direct comparisons were not performed.

Figure 8

Most Common Reasons for Apps Educators Would Not Download



Note: Error bars display one standard deviation above the mean. Lines above each bar indicate homogeneous subgroups. All reasons under the same bar are not statistically different from one another as determined by a Friedman's test using stepwise step-down comparisons. Reasons are organized in descending order by rank score as determined by the Friedman's test. Letters are assigned in order of the highest-ranked subgroup (a) to the lowest-ranked subgroup (f). *Starred reasons indicate that the reason's rank order as determined by the Friedman's test is different than the order for educators' overall reasons. Note that direct comparisons were not performed.

Part II Discussion

The present study identified 13 topics that educators cite as reasons for downloading an app when making app evaluations. Judgement was the overall most common reason, indicating that educators are making positive, negative, or neutral judgements of educational apps based on the information available to them. It is unsurprising that the majority (70%) of responses included a general judgement, as educators would necessarily need to make judgements to decide whether or not to download an app.

User Ratings and Reviews

As anticipated, user ratings and reviews were a common reason cited by educators, belonging to the second most common subgroup of reasons alongside educational applications/outcomes. Educators mentioned user reviews and ratings in approximately 59% of their responses, supporting findings from part I that user ratings had a strong impact on educators' app evaluations and corroborating previous work demonstrating that user ratings impact consumers' download decisions (Biviji et al., 2020; Burgers et al., 2016; Krishnan & Selvam, 2019). In fact, user reviews/ratings were mentioned significantly more frequently than educational benchmarks, suggesting that they may be more influential in educators' decisions than benchmarks. Unfortunately, ratings are not necessarily associated with educational quality (Dubé et al., 2020). Thus, to choose apps of high educational quality, educators should be encouraged to look for educational benchmarks rather than relying on ratings and reviews.

Rankings

In contrast to ratings and reviews, rankings were the least common reason mentioned by educators, and were mentioned in less than 1% of educators' responses. Together with part I, which revealed the surprising result that educators were more likely to download bottom-ranking

than top-ranking apps, this result may indicate that rankings are not an important consideration for educators. Instead, educators' apparent preference for bottom-ranking apps could be explained by other considerations. For example, despite the random assignment of images, educators may have preferred the visual characteristics of the bottom-ranking app pages. Indeed, visuals formed the third most frequently mentioned reason; they were mentioned in approximately 20% of responses - more than benchmarks or buzzwords. Although visual complexity of the apps in this study was controlled based on evidence that consumers have preferences for different levels of visual complexity (Wang & Lin, 2019), educators may consider other aspects of visual design as well. For example, "colorful" was a word that educators frequently mentioned. The complexity measure used in this study could account for the number of colours used in an app page, but not the perceived attractiveness of those colours. Wang and Lin (2019) identified that young children appreciate "colourfulness" in e-learning web pages, as defined by both the number of colours and their attractiveness. Educators may therefore choose apps that they perceive to have an attractive colour composition, expecting that such designs would be appreciated by students. Further investigation into educators' visual design preferences for educational apps is warranted.

Educational Benchmarks

Educational benchmarks were a common reason cited by educators, mentioned in approximately 15% of responses. However, in contrast to part I, which revealed that educators preferred educational apps including benchmarks, there was no evidence that educators mentioned benchmarks more frequently than buzzwords. In addition, benchmarks were significantly less common overall than judgement, academic applications/outcomes, and visuals, and there was no evidence that benchmarks were more common than more information,

cost/benefit, theme, and suitability. This result could indicate that benchmarks are not as influential in educators' decisions as previous research (Montazami et al., 2022a) and part I may suggest. Yet, educators frequently discussed academic applications/outcomes, showing that they are considering the educational potential of each app. Particularly for apps that educators said they would download, academic applications/outcomes was among the most common reasons. Clearly, an app's educational potential is important to educators. Why educational benchmarks were not a more frequent reason for whether to download an app is unclear.

Educators frequently described their reasons for downloading an app in terms of their own judgements and discussion of potential academic applications/outcomes. It is possible that educational benchmarks informed these judgements even if they were not mentioned specifically. Furthermore, educators may simply have used different terminology to express what they liked about an app. For example, educators mentioned suitability in approximately 10% of responses. Suitability is closely related to scaffolding, a benchmark which can include supports that tailor level of difficulty to a child's needs (Hirsh-Pasek et al., 2015).

The results could also suggest that educators have difficulty distinguishing between terms used to describe benchmarks and buzzwords. If educators are not fully proficient in the language of benchmarks, they may demonstrate a preference for benchmark apps in part 1 without being aware of the reasoning behind their preferences and less likely to mention them in part 2. Montazami and colleagues (2022b) previously noted that educators used the buzzword "engaging" to describe benchmark apps. Thus, educators may have mentioned buzzwords and benchmarks indiscriminately, using buzzwords to describe benchmark apps and vice versa. Indeed, separate examination of responses for benchmark and buzzword apps in the present study revealed that educators did occasionally mention buzzwords in their responses for why

they would download benchmark apps. However, benchmarks were mentioned more frequently than buzzwords for benchmark apps, while buzzwords were mentioned more frequently than benchmarks for buzzword apps, suggesting some distinction in how these terms were used.

Limitations

It is important to note that the present study did not directly compare the frequency of reasons between different types of apps. Doing so would have involved a very large number of comparisons, greatly limiting statistical power. As such, it was not possible to determine whether certain reasons were more common for benchmark apps than buzzword apps. Future research could focus on a limited number of the reasons identified in the present study for further investigation. Narrowing the scope in this manner would facilitate direct comparisons between benchmark and buzzword apps, improving our understanding of educators' apparent preference for benchmark apps. Investigating differences in reasons between apps educators would versus would not download could provide further insight into educators' app preferences.

The word frequency analysis techniques employed in this study present some additional limitations. First, our approach to classifying words into a single category may have resulted in some responses being incorrectly classified. For example, the word "space" was classified in the "theme" category because this was the context in which it was most frequently used. However, if an educator used the word space in another context (e.g., "the app space") this response would still be classified as mentioning app theme. Second, only words with a frequency greater than 10 were categorized and considered in analyses. As a result, some responses may not have been classified in categories that they should have been. For example, if educators mentioned an app theme with a frequency less than 10 (e.g., "safari" had a frequency of 5) their response would not be classified as mentioning app theme. Because we categorized variations of the same word

individually based on context, even variations of the same word were not classified if their frequency was less than 10 (e.g., colorful was classified while color was not). Finally, this approach does not provide information about whether reasons were discussed in a positive or negative light. We might expect educators to discuss positive reasons for apps they would download, and negative reasons for apps they would not, but this is not guaranteed.

Despite its limitations, the word frequency analysis technique employed in this study remains a useful way to identify common topics efficiently in a large amount of text (over 50,000 words) while maintaining the original context in which various terms were used. Other text mining approaches classify text automatically (i.e., topic modeling), but may not create the themes that are expected by the researcher (van Loon, 2022). Thus, such approaches were deemed inappropriate because they could not classify words into categories that were defined and manipulated in the present study, such as educational buzzwords and benchmarks. As such, our analyses provide a useful overview of the common topics educators cite as reasons for whether they would download an app. However, comparisons of frequencies should be interpreted with caution. Future research should take steps to validate the word categories identified in the present study, such as by comparing classification of responses using the word frequency approach to manual classification by human coders.

Overall Discussion

Ratings

Taken together, the results of this study clearly demonstrate that user ratings and reviews are important considerations in educators' decisions to download an educational app. Yet, it is unclear how useful they are for effectively evaluating the educational quality of an app. Dubé and colleagues (2020) found no evidence that star ratings are associated with whether an app's

description mentions educational benchmarks, suggesting that star ratings may not be a useful indicator of educational quality. However, an app's description in the App Store may differ from the true contents of the app. To better understand the extent to which user ratings are associated with educational quality, future research could compare an app's user ratings to the ratings of experts who have tested and evaluated the educational quality of the app. Evaluating the extent to which user ratings are associated with app quality will provide further support for whether interventions are needed to improve the selection of educational apps used in schools.

Furthermore, app reviews do sometimes include information about the educational contents of an app (Singh & Suri, 2022), which educators could use to help assess educational quality. Although the reviews in the present study did not include this information, future studies could explore whether educators are more influenced by app reviews that do mention educational content (e.g., educational benchmarks). If educators value this information in app reviews, it may be appropriate to focus intervention efforts on improving app store review systems so that they are more useful for educators rather than discouraging educators from relying on reviews.

Educational Benchmarks

The present study also highlights the need for further exploration on the role of educational benchmarks in educators' app selection decisions. While the findings support previous research showing educators prefer apps whose descriptions include educational benchmarks (Montazami et al., 2022), they also call into question the extent to which educators identify and acknowledge benchmarks in their decisions. Future research should investigate educators' proficiency in distinguishing benchmarks from buzzwords, clarifying whether they have sufficient knowledge to reliably use benchmarks to choose high-quality educational apps. Examining whether educators' existing Technological, Pedagogical, and Content Knowledge is

related to their preferences for benchmark apps would further provide insight into whether app selection could be improved by building educators' knowledge of technology and pedagogy.

Limitations

In addition to the limitations specific to parts I and II, this study includes some additional limitations that warrant investigation in future research. First, the present study addresses Montazami and colleagues' (2022a) concerns regarding ecological validity by considering how ratings, rankings, and other considerations impact educators' app evaluations. Nevertheless, the issue of ecological validity warrants further consideration. App pages were manipulated to control for price, length, and other factors that might influence educators' decisions in real life. Educators' reasons for downloading an app were consequently constrained by the content of the app pages that we manipulated. If asked to describe their reasoning for downloading real apps, educators may discuss additional topics that were not included in the present study. For example, in-app advertising has been identified as an impediment to students' learning (Falloon, 2013), but was not mentioned in any of the app pages for the present study. Further, this study focused only on one small part of the app selection process. Educators did not have the opportunity to test out apps or consult external sources for information. Indeed, many educators mentioned the need for more information about an app as a consideration for whether to download it. Larkin (2013) found that app descriptions often do not align with the app's actual contents, making it necessary for educators to test out apps after downloading them. Urquhart and colleagues (2023) explored parents' app evaluations by asking them to test a series of real educational apps and rate them based on their educational quality. Future research could explore educators' app selection process by asking them to select, download, and test real apps.

A final limitation of the present study pertains to its sample, which was composed of primarily white (87%) women (67%), and was recruited entirely online, limiting its generalizability. Notably, educators who opt to participate in online surveys may be more comfortable with technology, which could impact their approach to app selection. Future studies should recruit a diverse sample of educators through non-electronic means to explore a larger range of perspectives. It would also be worthwhile to consider individual differences in educators' app preferences, such as by exploring how educators' prior knowledge and experience with technology impacts their choices and reasoning. Furthermore, contextual factors, such as the location and type of school where educators teach, likely impact educators' choices and warrant further exploration.

Conclusions

Educators are ultimately responsible for providing opportunities for meaningful learning within a school environment. As such, it is essential that any technology they bring into the classroom can facilitate learning. If educators are to use mobile devices, as is common in elementary schools (Pearson, 2015), they must therefore choose high-quality educational apps.

This thesis aimed to understand what information impacts educators' evaluation and selection of educational apps. Part I determined that educational benchmarks, user ratings, and app store rankings all impact educators' app decisions. User ratings were particularly impactful (i.e., had the largest effect sizes) with educators having a strong preference for positively-rated apps, while the impact of app store rankings was less clear. Part II identified 13 topics that educators cite as reasons for whether they would download an educational app. Ratings/reviews were one of the most common reasons, while app rankings were least common. Educational benchmarks were a common reason; however, they were no more common than buzzwords.

Some additional reasons were identified, warranting future study. Notably, educators more frequently cited visuals as a reason for whether to download an app than benchmarks or buzzwords, highlighting a need for further investigation of how educators' preferences for visual characteristics of an app (e.g., colour) may impact their decisions.

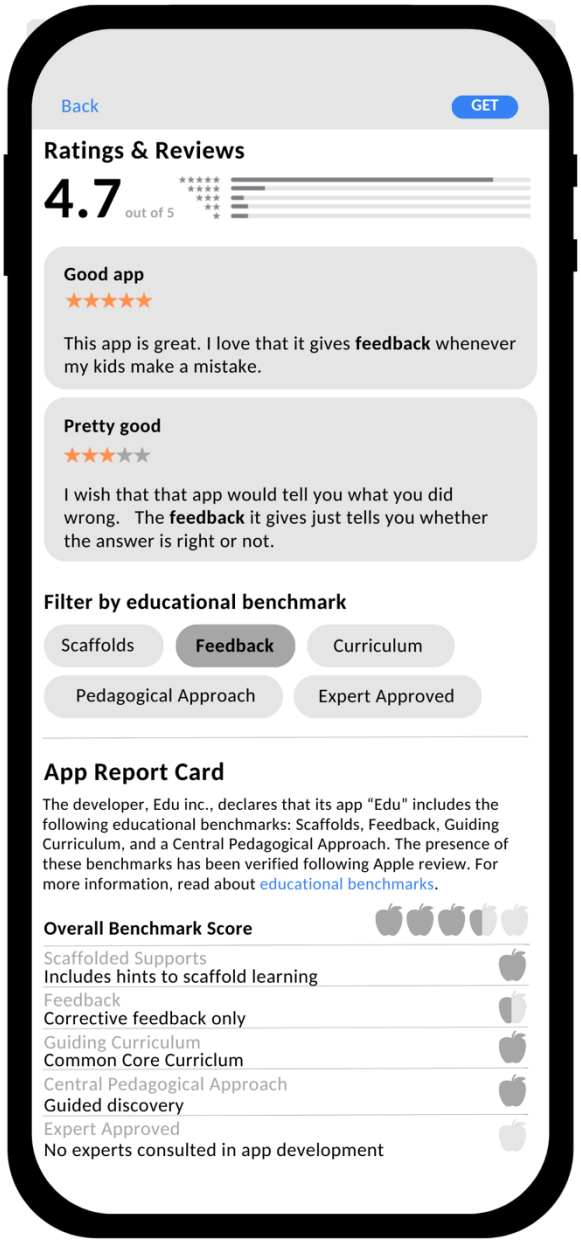
The present research supports previous findings (Montazami et al., 2022a) that educators value five educational benchmarks when choosing apps. It is recommended that more app developers include benchmarks in their apps and app store pages, as they are valued by educators and will hopefully encourage development of higher-quality apps. However, educators' preference for educational benchmarks appears to be overshadowed by their preference for apps with high user ratings. While educators should ideally use their own Technological, Pedagogical, and Content knowledge to evaluate which apps are appropriate for use in the classroom, our results suggest that they are also strongly influenced by others' judgements of app quality. Although educators clearly want to choose apps of high educational quality, user ratings and reviews are a more common reason for whether they would download an app than benchmarks of educational quality. This finding has implications for researchers, who we encourage to develop interventions and training programs that can help educators build confidence in their knowledge and use it to choose high-quality apps (cf., user ratings).

Further, educators may benefit from improvements to existing app store review systems that would make them more useful for judging app quality. App stores and the companies that run them make considerable profits off educational apps that are purchased by educators and schools using tax dollars. It should be incumbent on these companies to provide transparent and meaningful information about the educational products they sell. If app stores were to introduce a system whereby developers report the presence of benchmarks and consumers review apps

based on their benchmarks (see Figure 9), it would be easier for educators to rely on ratings to make effective app decisions. Clearly, app store user ratings are a key factor in educators' educational app selection; now, it is up to app stores to make user ratings meaningful.

Figure 9

Proposed Educational Quality App Store Review Criteria



References

Apple Inc. (2010, March 5). *IPad available in US on April 3*. Apple Newsroom. https://www.apple.com/newsroom/2010/03/05iPad-Available-in-US-on-April-3/

Apple Inc. (2023a). *K–12 education*. Apple. https://www.apple.com/ca/education/k12/

Apple Inc. (2023b, June 5). *App store review guidelines*. Apple Developer. https://developer.apple.com/app-store/review/guidelines/

ASCD. (2019). *The 2019 K-12 Digital Content Report*.

Bakker, M., & Wicherts, J. M. (2014). Outlier removal, sum scores, and the inflation of the Type I error rate in independent samples t tests: The power of alternatives and recommendations. *Psychological Methods, 19*(3), 409–427. https://doi.org/10.1037/met0000014

Bebell, D., & Pedulla, J. (2015). A quantitative investigation into the impacts of 1:1 iPads on early learner's ELA and math achievement. *Journal of Information Technology Education: Innovations in Practice*, *14*, 191–215. https://doi.org/10.28945/2175

Biviji, R., Vest, J. R., Dixon, B. E., Cullen, T., & Harle, C. A. (2020). Factors related to user ratings and user downloads of mobile apps for maternal and infant health: Cross-sectional study. *JMIR MHealth and UHealth*, *8*(1), e15663. https://doi.org/10.2196/15663

Blacker, A. (2022, February 1). *App store rank explained: What is an app's Apple & Google store ranking and what impacts it?* https://blog.apptopia.com/app-store-rank-explained-what-is-an-apps-apple-google-store-ranking-and-what-impacts-it

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42. https://doi.org/10.3102/0013189X018001032

Bruckman, A. (1998). Community support for constructionist learning. *Computer Supported Cooperative Work (CSCW)*, *7*(1), 47–86. https://doi.org/10.1023/A:1008684120893

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, *31*, 21–32.

Burgers, C., Eden, A., de Jong, R., & Buningh, S. (2016). Rousing reviews and instigative images: The impact of online reviews and visual design characteristics on app downloads. *Mobile Media & Communication*, *4*(3), 327–346. https://doi.org/10.1177/2050157916639348

Cai, Z., Mao, P., Wang, D., He, J., Chen, X., & Fan, X. (2022). Effects of scaffolding in digital game-based learning on student's achievement: A three-level meta-analysis. *Educational Psychology Review*, *34*(2), 537–574. https://doi.org/10.1007/s10648-021-09655-0

Callaghan, M. N., & Reich, S. M. (2018). Are educational preschool apps designed to teach? An analysis of the app market. *Learning, Media and Technology*, *43*(3), 280–293. https://doi.org/10.1080/17439884.2018.1498355

Camilleri, A., & Camilleri, M. A. (2019). Mobile learning via educational apps: An interpretative study. In N. G. Shun-Wing, T. S. Fun, & Y. Shi (Eds.), *Proceedings of the 2019 5th International Conference on Education and Training Technologies* (pp. 88–92). ACM. https://doi.org/10.1145/3337682.3337687

Campbell, G., & Skillings, J. H. (1985). *Nonparametric Stepwise Multiple Comparison Procedures. Journal of the American Statistical Association, 80*(392), 998–1003. https://doi.org/10.2307/2288566

Campbell, J., & Thompson, V. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, *44*. https://doi.org/10.3758/s13428-012-0186-0

Cao, Y., Proctor, R. W., Ding, Y., Duffy, V. G., Zhang, Y., & Zhang, X. (2022). Is an anthropomorphic app icon more attractive than a non-anthropomorphic one? A case study using multimodal measurement. *International Journal of Mobile Communications*, *20*(4), 419–439. https://doi.org/10.1504/IJMC.2022.123789

Cao, Y., Zhang, Y., Ding, Y., Duffy, V. G., & Zhang, X. (2021). Is an anthropomorphic app icon more attractive? Evidence from neuroergonomomics. *Applied Ergonomics*, *97*, 103545. https://doi.org/10.1016/j.apergo.2021.103545

Carare, O. (2012). The impact of bestseller rank on demand: Evidence from the app market*. *International Economic Review*, *53*(3), 717–742. https://doi.org/10.1111/j.1468-2354.2012.00698.x

Carr, J. M. (2012). Does math achievement h'APP'en when iPads and game-based learning are incorporated into fifth-grade mathematics instruction? *Journal of Information Technology Education: Research*, *11*(1), 269–286.

Cayton-Hodges, G. A., Feng, G., & Pan, X. (2015). Tablet-based math assessment: What can we learn from math apps? *Journal of Educational Technology & Society*, *18*(2), 3–20.

Cherner, T., Dix, J., & Lee, C. (2014). Cleaning up that mess: A framework for classifying educational apps. *Contemporary Issues in Technology and Teacher Education*, *14*(2), 158–193.

Common Sense Education. (n.d.). *Best Games, Apps, and Sites for Personalized Learning | Common Sense Education*. Retrieved August 18, 2023, from https://www.commonsense.org/education/lists/best-games-apps-and-sites-for-personalized-learning

Crescenzi-Lanna, L., & Grané-Oró, M. (2016). An analysis of the interaction design of the best educational apps for children aged zero to eight. *Comunicar: Revista Científica de Comunicacíon y Educacíon, 24*(46), 77–85. https://doi.org/10.3916/C46-2016-08

Dewey, J. (1938). *Experience and education.* Simon and Schuster.

Dubé, A. K., Alam, S. S., Xu, C., Wen, R., & Kacmaz, G. (2019). Tablets as elementary mathematics education tools: Are they effective and why. In K. M. Robinson, H. P. Osana, & D. Kotsopoulos (Eds.), *Mathematical Learning and Cognition in Early Childhood: Integrating Interdisciplinary Research into Practice* (pp. 223–248). Springer International Publishing. https://doi.org/10.1007/978-3-030-12895-1_13

Dubé, A. K., Kacmaz, G., Wen, R., Alam, S. S., & Xu, C. (2020). Identifying quality educational apps: Lessons from 'top' mathematics apps in the Apple App store. *Education and Information Technologies*, *25*(6), 5389–5404. https://doi.org/10.1007/s10639-020-10234-z

Dündar, H., & Akçayır, M. (2014). Implementing tablet PCs in schools: Students' attitudes and opinions. *Computers in Human Behavior*, *32*, 40–46. https://doi.org/10.1016/j.chb.2013.11.020

Fabian, K., Topping, K. J., & Barron, I. G. (2016). Mobile technology and mathematics: Effects on students' attitudes, engagement, and achievement. *Journal of Computers in Education*, *3*(1), 77–104. https://doi.org/10.1007/s40692-015-0048-8

Falloon, G. (2013). Young students using iPads: App design and content influences on their learning pathways. *Computers & Education*, *68*, 505–521. https://doi.org/10.1016/j.compedu.2013.06.006

Falloon, G. (2017). Mobile devices and apps as scaffolds to science learning in the primary classroom. *Journal of Science Education and Technology*, *26*(6), 613–628. https://doi.org/10.1007/s10956-017-9702-4

Farber, M. (2015, September 8). *Hands-on apps for diverse learners*. Edutopia. https://www.edutopia.org/blog/hands-on-apps-diverse-learners-matthew-farber

Fernández-Batanero, J.-M., Román-Graván, P., Reyes-Rebollo, M.-M., & Montenegro-Rueda, M. (2021). Impact of educational technology on teacher stress and anxiety: A literature review. *International Journal of Environmental Research and Public Health*, *18*(2), Article 2. https://doi.org/10.3390/ijerph18020548

Furner, C. P., & Zinko, R. (2018). Willingness to pay and disposition toward paying for apps: The influence of application reviews. *International Journal of E-Services and Mobile Applications*, *10*(1), 13–33. https://doi.org/10.4018/IJESMA.2018010102

Guru Technolabs, G. (2023, August 3). Key features every educational app should have. *Medium*. https://medium.com/@GuruTechnolabs/key-features-every-educational-app-should-have-3895b33db7f4

Hair, J., Black, W., Babin, B., & Anderson, R. (2010). *Multivariate data analysis* (7th Edition). Pearson Prentice Hall.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest*, *16*(1), 3–34. https://doi.org/10.1177/1529100615569721

IBM. (2022). *IBM SPSS Statistics Algorithms.* https://www.ibm.com/docs/en/SSLVMB_29.0.0/pdf/IBM_SPSS_Statistics_Algorithms.pdf

Javora, O., Hannemann, T., Stárková, T., Volná, K., & Brom, C. (2019). Children like it more but don't learn more: Effects of esthetic visual design in educational games. *British Journal of Educational Technology, 50*(4), 1942–1960. https://doi.org/10.1111/bjet.12701

Jylhä, H., & Hamari, J. (2019). An icon that everyone wants to click: How perceived aesthetic qualities predict app icon successfulness. *International Journal of Human-Computer Studies*, *130*, 73–85. https://doi.org/10.1016/j.ijhcs.2019.04.004

Kafai, Y. B., & Resnick, M. (1996). *Constructionism in practice: Designing, thinking, and learning in a digital world*. Routledge.

Karch, J. D. (2023). Outliers may not be automatically removed. *Journal of Experimental Psychology: General*, *152*(6), 1735–1753. https://doi.org/10.1037/xge0001357

Kebritchi, M., & Hirumi, A. (2008). Examining the pedagogical foundations of modern educational computer games. *Computers & Education*, *51*(4), 1729–1743. https://doi.org/10.1016/j.compedu.2008.05.004

Kim, J. (2023, August 29). *Top features in education apps*. GoodFirms. https://www.goodfirms.co/blog/features-of-education-app

Kim, J., Gilbert, J., Yu, Q., & Gale, C. (2021). Measures matter: A meta-analysis of the effects of educational apps on preschool to grade 3 children's literacy and math skills. *AERA Open*, *7*. https://doi.org/10.1177/23328584211004183

Joyce, B. R., Weil, M., & Showers, B. (1992). *Models of teaching* (4th ed.). Allyn & Bacon.

Kolak, J., Norgate, Sarah. H., Monaghan, P., & Taylor, G. (2021). Developing evaluation tools for assessing the educational potential of apps for preschool children in the UK. *Journal of Children and Media*, *15*(3), 410–430. https://doi.org/10.1080/17482798.2020.1844776

Krishnan, G., & Selvam, G. (2019). Factors influencing the download of mobile health apps: Content review-led regression analysis. *Health Policy and Technology*, *8*(4), 356–364. https://doi.org/10.1016/j.hlpt.2019.09.001

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Larkin, K. (2013). Mathematics Education: Is There an App for That? In V. Steinle, L. Ball, & C. Bardini (Eds.), *Mathematics education: Yesterday, today and tomorrow: 36th Annual Conference of the Mathematics Education Research Group of Australasia (MERGA).* https://eric.ed.gov/?id=ED572942

LASER Institute. (2021). *Text Mining in STEM Ed Research*. https://github.com/laser-institute/text-mining

Lee, C.-Y., & Cherner, T. S. (2015). A comprehensive evaluation rubric for assessing instructional apps. *Journal of Information Technology Education: Research*, *14*, 021–053. https://doi.org/10.28945/2097

List, A., Lee, H. Y., Du, H., Campos Oaxaca, G. S., Lyu, B., Falcon, A. L., & Lin, C.-J. (2022). Preservice teachers' recognition of source and content bias in educational application (app) reviews. *Computers in Human Behavior*, *134*, 107297. https://doi.org/10.1016/j.chb.2022.107297

Maalej, W., Kurtanović, Z., Nabil, H., & Stanik, C. (2016). On the automatic classification of app reviews. *Requirements Engineering*, *21*(3), 311–331. https://doi.org/10.1007/s00766-016-0251-9

Manchanda, A. (2022, August 24). 13 must-have features of successful educational apps. *Insights - Web and Mobile Development Services and Solutions*. https://www.netsolutions.com/insights/effective-features-of-best-educational-apps/

Mayring, P. (2015). Qualitative Content Analysis: Theoretical Background and Procedures. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to Qualitative Research in Mathematics Education: Examples of Methodology and Methods* (pp. 365–380). Springer Netherlands. https://doi.org/10.1007/978-94-017-9181-6_13

McIlroy, S., Ali, N., Khalid, H., & E. Hassan, A. (2016). Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews. *Empirical Software Engineering*, *21*(3), 1067–1106. https://doi.org/10.1007/s10664-015-9375-7

Meyer, M., Zosh, J. M., McLaren, C., Robb, M., McCaffery, H., Golinkoff, R. M., Hirsh-Pasek, K., & Radesky, J. (2021). How educational are "educational" apps for young children? App store content analysis using the Four Pillars of Learning framework. *Journal of Children and Media*, *15*(4), 526–548. https://doi.org/10.1080/17482798.2021.1882516

Meyers, L. S., Gamst, G., & Guarino, A. J. (2016). *Applied multivariate research: Design and interpretation*. SAGE Publications.

Mishra, P., & Koehler, M. J. (2008). Introducing Technological Pedagogical Content Knowledge. In AACTE Committee on Innovation and Technology (Ed.), *Handbook of Technological Pedagogical Content Knowledge (TPCK) for Educators* (pp. 3–29). Routledge.

Montazami, A., Pearson, H. A., Dubé, A. K., Kacmaz, G., Wen, R., & Alam, S. S. (2022a). Why this app? How educators choose a good educational app. Computers & Education, 184, 104513. https://doi.org/10.1016/j.compedu.2022.104513

Montazami, A., Pearson, H. A., Dubé, A. K., Kacmaz, G., Wen, R., & Alam, S. S. (2022b). Why this app? How parents choose good educational apps from app stores. *British Journal of Educational Technology*, *53*(6), 1766–1792. https://doi.org/10.1111/bjet.13213

Moyer-Packenham, P. S., Lommatsch, C. W., Litster, K., Ashby, J., Bullock, E. K., Roxburgh, A. L., Shumway, J. F., Speed, E., Covington, B., Hartmann, C., Clarke-Midura, J., Skaria, J., Westenskow, A., MacDonald, B., Symanzik, J., & Jordan, K. (2018). How design features in digital math games support learning and mathematics connections. *Computers in Human Behavior*, *91*, 316–332. https://doi.org/10.1016/j.chb.2018.09.036

Olson, C. L. (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, *83*(4), 579–586. https://doi.org/10.1037/0033-2909.83.4.579

OpenAI. (2022). *ChatGPT* (Version 3.5) [Computer software]. https://openai.com/chatgpt

Ormrod, J. (1995). *Educational psychology: Principles and applications*. Prentice Hall.

Pagano, D., & Maalej, W. (2013). User feedback in the appstore: An empirical study. *2013 21st IEEE International Requirements Engineering Conference (RE)*, 125–134. https://doi.org/10.1109/RE.2013.6636712

Papadakis, S., & Kalogiannakis, M. (2017). Mobile educational applications for children. What educators and parents need to know. *International Journal of Mobile Learning and Organisation*, *11*, 1. https://doi.org/10.1504/IJMLO.2017.10003925

Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2017). Designing and creating an educational app rubric for preschool teachers. *Education and Information Technologies*, *22*(6), 3147–3165. https://doi.org/10.1007/s10639-017-9579-0

Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2018). Educational apps from the Android Google Play for Greek preschoolers: A systematic review. *Computers & Education*, *116*, 139–160. https://doi.org/10.1016/j.compedu.2017.09.007

Pearson. (2015). *Pearson student mobile device survey 2015*. https://www.pearson.com/content/dam/one-dot-com/one-dot-com/us/en/pearson-ed/downloads/2015-Pearson-Student-Mobile-Device-Survey-Grades-4-12.pdf

Pearson, H. A., Montazami, A., & Dubé, A. K. (2023). Why this app: Can a video-based intervention help parents identify quality educational apps? *British Journal of Educational Technology*, *54*(3), 712–733. https://doi.org/10.1111/bjet.13284

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Petty, G. (2009). *Evidence-based teaching: A practical approach* (Second edition). Nelson Thornes. https://cir.nii.ac.jp/crid/1130000794532163584

Pocket Gamer. (2024). *App store metrics*. https://www.pocketgamer.biz/metrics/app-store/

R Core Team. (2023). *R: A language and environment for statistical computing* (4.3.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rosell-Aguilar, F. (2017). State of the app: A taxonomy and framework for evaluating language learning mobile applications. *CALICO Journal*, *34*(2), 243–258.

Scheff, S. W. (2016). Chapter 8—Nonparametric statistics. In S. W. Scheff (Ed.), *Fundamental statistical principles for the neurobiologist* (pp. 157–182). Academic Press. https://doi.org/10.1016/B978-0-12-804753-8.00008-7

Shahjad, & Mustafa, K. (2022). A systematic literature review on learning apps evaluation. *Journal of Information Technology Education: Research*, *21*, 663–700. https://doi.org/10.28945/5042

Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in R. *JOSS, 1*(3). https://doi.org/10.21105/joss.00037

Singh, Y., & Suri, P. K. (2022). An empirical analysis of mobile learning app usage experience. *Technology in Society*, *68*, 101929. https://doi.org/10.1016/j.techsoc.2022.101929

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th edition*)*. Allyn & Bacon/Pearson Education.

Takeuchi, L. M., & Vaala, S. (2014). *Level up learning: A national survey on teaching with digital games*. The Joan Ganz Cooney Center at Sesame Workshop.

Tärning, B. (2018). Review of feedback in digital applications – Does the feedback they provide support learning? *Journal of Information Technology Education: Research*, *17*, 247–283. https://doi.org/10.28945/4104

Taylor, G., Kolak, J., Bent, E. M., & Monaghan, P. (2022). Selecting educational apps for preschool children: How useful are website app rating systems? *British Journal of Educational Technology*, *53*(5), 1262–1282. https://doi.org/10.1111/bjet.13199

Taylor, G., Kolak, J., Norgate, S. H., & Monaghan, P. (2022). Assessing the educational potential and language content of touchscreen apps for preschool children. *Computers and Education Open*, *3*, 100102. https://doi.org/10.1016/j.caeo.2022.100102

Tongdee, E., & Markowitz, O. (2018). Mobile app rankings in dermatology. *Cutis*, *102*(4), 252–256.

Urquhart, N., Lee, J., & Wood, E. (2023). Get that app!: Examining parental evaluations of numeracy apps. *Journal of Research in Childhood Education*, 1–15. https://doi.org/10.1080/02568543.2023.2260433

Vaala, S., Ly, A., & Levine, M. H. (2015). Getting a read on the app stores: A market scan and analysis of children's literacy apps. Full Report. In *Joan Ganz Cooney Center at Sesame Workshop*. Joan Ganz Cooney Center at Sesame Workshop. https://eric.ed.gov/?id=ED574396

van Loon, A. (2022). Three families of automated text analysis. *Social Science Research, 108*, 102798. https://doi.org/10.1016/j.ssresearch.2022.102798

Walker, H. (2011). Evaluating the effectiveness of apps for mobile devices. *Journal of Special Education Technology*, *26*(4), 59–63.

Walz, A. (2015, May 27). *Deconstructing the App Store rankings formula with a little mad science*. Moz. https://moz.com/blog/app-store-rankings-formula-deconstructed-in-5-mad-science-experiments

Wang, M., & Li, X. (2017). Effects of the aesthetic design of icons on app downloads: Evidence from an android market. *Electronic Commerce Research*, *17*(1), 83–102. https://doi.org/10.1007/s10660-016-9245-4

Wang, H.-F., & Lin, C.-H. (2019). An investigation into visual complexity and aesthetic preference to facilitate the creation of more appropriate learning analytics systems for children. *Computers in Human Behavior, 92*, 706–715. https://doi.org/10.1016/j.chb.2018.05.032

Wilcox, R. (1998). How many discoveries have been lost by ignoring modem statistical methods? *American Psychologist*, *53*(3), 300–314.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., Miller, E., & Smith, D. (2023). *haven: Import and Export "SPSS", "Stata" and "SAS" Files* (2.5.4) [Computer software]. https://CRAN.R-project.org/package=haven

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.03087

Wisniewski, H., Liu, G., Henson, P., Vaidyam, A., Hajratalli, N. K., Onnela, J.-P., & Torous, J. (2019). Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. *BMJ Ment Health*, *22*(1), 4–9. https://doi.org/10.1136/ebmental-2018-300069

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving*. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Zydney, J. M., & Warner, Z. (2016). Mobile apps for science learning: Review of research. *Computers & Education*, *94*, 1–17. https://doi.org/10.1016/j.compedu.2015.11.001

Appendix A

Manipulated Information of Fake Educational App Pages

	Condition	Text Description	Reviews
Math Safari App type Rating Ranking	 Buzzword Negative (2.3) Bottom (86)	In our app, kids develop their math understanding as they are taken on a safari adventure. Kids will love practicing math in Math Safari! ** Math Safari allows kids to practice and learn math content in a completely interactive way. ** Math Safari uses multi-media to bring math to life for kids. ** Kids can learn new math concepts in a fully hands-on way by participating in a variety of fun challenges, games, and lessons. ** Exciting animations and sound effects help to ensure that using Math Safari will be an engaging learning experience for kids. ** Children can personalize their learning by creating their own characters.	**Awful** This is my kids' least favourite app (and mine too). **Just the worst** Please don't download this app. It's awful.
Dog Math App type Rating Ranking	 Buzzword Negative (2.8) Middle (43)	Use math to help Spot the dog navigate daily life at the park. Dog Math will make kids excited to learn math! ** Fun challenges and activities make Dog Math a fully interactive learning experience. ** With a variety of exciting sights and sounds, Dog Math uses multi-media to make math fun. ** All of the activities in Dog Math are totally hands-on, making our app an excellent way for kids to practice their math skills. ** Dog Math is so engaging that your kids will never want to stop practicing math!	**Regrets** Why did I ever download this terrible app for my kids? It was such a waste of time. **What a nightmare** This is an absolute nightmare of an app.

		** Our personalized math app will help your child excel in math like they never have before.	
Animal Math App type Rating Ranking	 Buzzword Negative (1.1) Top (9)	Kids will save the animals as they learn and improve their math skills. They'll love learning with Animal Math! ** Interactive features make Animal Math an enjoyable and effective way for kids to learn math. ** In Animal Math a variety of multi-media features come together to make math learning easy and fun. ** With a variety of hands on-activities, Animal Math helps kids master the foundational math skills they need to succeed. ** The engaging design of Animal Math keeps kids coming back for more and more math learning fun. ** Kids personalize their learning in Animal Math by choosing their own fun characters and colourful backgrounds.	**Ughhh** This app was a complete waste of my kids' time. **Trash** This app is trash. Don't waste your time.
Time for Math App type Rating Ranking	 Buzzword Neutral (3.6) Bottom (88)	In Time for Math, kids are taken on an incredible journey through math learning. Our app will make kids wish it was always time for math! ** Filled with interactive activities that make learning math enjoyable and rewarding. ** Games and exercises are enhanced with multi-media content such as videos and animations. ** Time for Math is designed to be fun and hands-on, making it the perfect tool for helping kids learn math. ** Time for Math is designed to make math learning engaging and help students practice their skills. ** Cool puzzles and fun challenges make Time for Math a great way to personalize your child's learning.	**Okay** This isn't my kids' favourite math learning app, but it isn't their least favourite either. **Nothing special** There are probably better options out there, but this isn't a bad app.

Math is Sweet App type Rating Ranking	 Buzzword Neutral (3.4) Middle (49)	Kids will develop a taste for math in Math is Sweet. Practice math skills with a series of delectable math puzzles! ** Math is Sweet is filled with interactive games and exercises that help kids understand and retain math concepts. ** Math is Sweet allows children to learn a variety of math topics using an exciting multi-media interface. ** Improve math knowledge by practicing your skills with a variety of hands-on problems to solve. ** Features engaging math challenges and games that kids can play over and over again. ** With personalized learning in Math is Sweet, math education will never be the same.	**Average** I'd say my kids' experience using this app has been pretty average. **Can't complain** There's better math apps, but I can't complain.
Math: Let's Play! App type Rating Ranking	 Buzzword Neutral (3.8) Top (3)	Develop a love for math with Math: Let's Play! This app makes it easy and fun for kids to learn and practice math. ** Our math app is packed with interactive activities that make learning math fun for kids. ** Math: Let's Play! allows kids to explore math on an exciting multi-media platform. ** Kids can drag and drop to solve math problems, making it a totally hands-on learning experience. ** This app keeps kids engaged while they learn a variety of important math concepts. ** Math: Let's Play! is a great way for your child to personalize their math learning and skills development.	**Pretty basic** This app is a bit basic, but it does the job I need it to do. **Neutral thoughts** This app isn't terrible, but it's not the best. It's fine for what it is.
Math Superstar App type Rating Ranking	 Buzzword Positive (4.6) Bottom (89)	In our app, kids will develop their math skills on an exciting adventure. Kids will feel like a math superstar!	**Yessss** My kids never used to enjoy math, but now they can't

		** Kids will learn to love math due to Math Superstar's fully interactive interface and features. ** By incorporating both visual and auditory information, Math Superstar uses multi-media to enhance your child's learning experience. ** Hands-on elements make Math Superstar a great tool for helping kids learn math. ** Engaging sound effects and visuals make learning math a more enjoyable experience for kids. ** See how Math Superstar's personalized learning experience can help your child develop their math skills.	get enough of this app! **10/10** This app is super fun. I love it so much.
Math Math Math! App type Rating Ranking	 Buzzword Positive (4.7) Middle (47)	In Math Math Math! kids will become math masters. Help a team of friendly monsters save the world through the power of math! ** Math Math Math! is designed to be interactive, making it the perfect tool for helping kids learn math. ** This app is loaded with multi-media content to help kids practice math concepts and develop their skills. ** Math Math Math! helps kids learn tough math concepts in a hands-on way. ** Using this app is the most engaging way for children to develop their math skills. ** With tons of fun activities to kids to learn from, Math Math Math! personalizes your child's math development.	**So good** I can always hear my kid giggling at her iPad when she plays this. I can tell she loves it. **Wow** What a great app! I am very happy.
Let's Play Math App type Rating Ranking	 Buzzword Positive (4.9) Top (8)	Let's Play Math will help your kids enjoy math and develop their math skills! Learn math through a variety of fun challenges.	**Total game-changer** This app has been a game-changer in our house. I've

		** With Let's Play Math, kids can learn math through interactive activities that keep them excited about learning. ** Kids will easily be immersed in their math learning journey thanks to a variety of multi-media learning features. ** Kids can build up their math skills by participating in a wide selection of hands-on activities. ** Let's Play Math is designed to be an engaging way for kids to become experts on a wide range of math topics. ** Your child will want to keep learning and improving with Let's Play Math's personalized math program.	never seen my kids so happy to learn math! **Great app** I'm so glad I downloaded this great app.
Adventures in Math App type Rating Ranking	 Benchmark Negative (2.6) Bottom (85)	Go on an adventure through many worlds as you develop your skills in math! Kids will develop a love for learning with Adventures in Math. ** Our math app breaks down complex problems into smaller, more manageable chunks. ** Be assured that your kids are receiving an education that is in line with the standard school curriculum. ** Certified teachers and educational experts played an important role in Adventures in Math's development. ** This app gives immediate feedback, stating why your child's answer is correct or not. ** The learning theory behind Adventures in Math's design is supported by scientific evidence.	**Kid hates it** I never leave reviews, but I just had to this time to say how much my kid hates this app. **Worst math app** This is the worst kids' math learning app I have ever seen.
Super Math App type Rating Ranking	 Benchmark Negative (1.4) Middle (44)	Super Math offers an exciting journey of math practice. Kids will love learning math through fun math games and challenges! ** Super Math provides hints and support for students as they progress	**Terrible** My kids aren't learning anything from this awful app. **Just no** This app is the

		through increasingly challenging math problems. ** Activities are designed to help kids master the math skills outlined in the standard school curriculum. ** Experienced educators and researchers from top universities were consulted in the development of Super Math. ** Instant feedback enhances learning by explaining incorrect answers. ** Our app is built on a foundation of an evidence-based learning theory, ensuring that kids have the best possible chance of mastering math concepts.	WOOOOORRST.
Math Champion App type Rating Ranking	 Benchmark Negative (2.9) Top (6)	It won't take long for kids to become a Math Champion in our app! Kids improve their math performance by practicing math in fun puzzles. ** Math Champion reinforces key math skills with math questions that adapt to your child. ** Our math app helps kids meet the expectations of the standard school curriculum. ** Math Champion's development team includes researchers and educators who have extensive experience in teaching math to kids. ** The feedback provided by the app helps kids to identify and correct their mistakes. ** Development of Math Champion was guided by a scientifically tested theory of learning.	**Complete disaster** My kids refuse to play this app. It's that bad! **If I had a time machine…** I would go back in time and never download this terrible app.
Space Math App type Rating Ranking	 Benchmark Neutral (3.8) Bottom (83)	Kids can travel in a spaceship across the galaxy as they learn and practice math. Our app makes math out of this world! ** Our math app provides consistent support and guidance, ensuring that	**It's fine** My kids don't hate learning math with this app, but they don't really love it either.

		kids have the necessary tools to succeed. ** Space Math is aligned with the standard school curriculum ensuring that your child is learning the skills they need to succeed. ** Developed in consultation with developmental experts and educators to ensure the best learning experience for kids. ** Our math app uses a variety of feedback techniques to help kids learn from their mistakes and grow. ** To best aid children's math learning, an evidence-based theory of learning was central to Space Math's design.	**Decent** This is a decent app, but it's nothing special.
Magical Math		Solve exciting puzzles while learning math in Magical Math. Our app will help kids understand the magic of math! ** Magical Math gradually increases in difficulty to keep kids challenged. ** Our math app is the perfect tool for reinforcing the math concepts covered in the standard school curriculum. ** Developed by certified experts in child development and math educators. ** Each problem is followed by targeted feedback to help kids understand and correct their mistakes. ** Our app is based in an evidence-based theory of learning to help kids understand and retain math concepts.	**Mid-tier app** Don't expect anything spectacular, but this app did what I needed it to do. **Pretty good** This is a pretty good app for learning math.
App type	Benchmark		
Rating	Neutral (3.6)		
Ranking	Middle (45)		
Math Party		Math Party helps kids love math by allowing them to practice in a variety of fun games and activities! Improve your skills with Math Party. ** Strengthens key math skills with math questions that adapt to your child.	**I don't know** One of my kids likes it, but the other doesn't so I don't know. **Alright** This app is fine. Don't
App type	Benchmark		
Rating	Neutral (3.8)		
Ranking	Top (4)		

		** Math Party covers the main topics outlined in the standard school curriculum. ** Our development team includes math experts and classroom teachers who have dedicated their careers to helping kids learn math. ** Provides corrective feedback to help guide kids towards the right answer when they make a mistake. ** Using an evidence-based learning theory to guide its teaching methods, this app is designed to help kids succeed in math.	really have much to say about it.
Math Winter Adventure App type Rating Ranking	 Benchmark Positive (4.9) Bottom (84)	Can you use your math skills to make it through the winter? Find out as you develop your understanding in Math Winter Adventure. ** Math Winter Adventure breaks math down into smaller parts to help children solve problems that are just beyond their reach. ** Math Winter Adventure is aligned with the standard school curriculum. ** Our development team includes researchers and experienced educators who have a deep understanding of how kids learn math. ** Kids receive regular corrective feedback as they progress through the activities in this app. ** Math Winter Adventure's teaching methods and design are based on a research-backed theory of learning.	**Awesome** My kids fight over who gets to play this app. **Great app** I'm so glad I downloaded this great app.
Magical Math Bird App type Rating Ranking	 Benchmark Positive (4.5) Middle (48)	Put your math skills to the test as you help the magical math bird on an amazing quest! Kids will love learning math with Magical Math Bird. ** In Magical Math Bird, children are provided with guidance and support as they work to improve their math skills.	**Love it** Even in just a few weeks I can already tell my kids are getting better at math thanks to this app.

		** Based on a standard school curriculum, kids will learn math skills at their grade level. ** Magical Math Bird benefits from valuable input from experienced teachers and educational researchers. ** Kids receive detailed feedback after each problem, helping them to understand why they got it right or wrong. ** To ensure that Magical Math Bird is effective, its design was informed by a learning theory that is supported by educational research.	**Excellent app** This game is sooooo awesome. I really don't need to say anything else.
Zombie Math App type Rating Ranking	 Benchmark Positive (4.5) Top (5)	Use your math skills to beat the brain-dead and save the world with Zombie Math! Kids will love practicing math with our app! ** Each lesson builds upon the skills learned in the previous one, helping kids to master math concepts. ** Lessons are designed to provide a comprehensive math education that aligns with the standard school curriculum. ** This app is created by a group of dedicated teachers and developmental experts who are passionate about helping kids succeed in math. ** Zombie Math provides quick feedback after every question to help kids monitor their progress and learn from their mistakes. ** Zombie Math's teaching methods are based on a research-based theory of learning.	**AMAZING** My kids are always asking to play this app. **Math-tastic (haha)** This app is math-tastic! What more can I say?

Appendix B

Categorized Words and Frequencies

Word	Frequency	Category	Word	Frequency	Category
reviews	1310	Reviews/Ratings	designed	22	Benchmarks
app	1105	Uninformative	numbers	21	Uninformative
good	499	Judgement	specific	21	Uninformative
children	323	Academic A/O	extremely	21	Judgement
students	319	Academic A/O	helpful	21	Judgement
download	303	Uninformative	standard	21	Benchmarks
engaging	231	Buzzwords	pupils	21	Academic A/O
math	222	Academic A/O	visuals	20	Visuals
kids	215	Academic A/O	number	20	Uninformative
great	208	Judgement	older	20	Suitability
ratings	198	Reviews/Ratings	chart	20	Ranking
try	197	More Information	cool	20	Judgement
fun	183	Judgement	friendly	20	Judgement
learning	172	Academic A/O	playing	20	Gamification
bad	171	Judgement	picture	19	Visuals
based	168	Uninformative	concept	19	Uninformative
graphics	158	Visuals	fairly	19	Uninformative
feedback	158	Benchmarks	suggest	19	Uninformative
curriculum	145	Benchmarks	expectations	19	Benchmarks
poor	131	Judgement	lessons	19	Academic A/O
apps	130	Uninformative	buy	18	Uninformative
time	111	Cost/Benefit	hard	18	Uninformative
free	104	Cost/Benefit	provided	18	Uninformative
rating	100	Reviews/Ratings	shot	18	Uninformative
game	100	Gamification	parents	18	Reviews/Ratings
basic	96	Judgement	confusing	18	Judgement
description	92	Written Information	sufficient	18	Judgement
low	92	Reviews/Ratings	theory	18	Benchmarks
average	90	Judgement	progress	18	Academic A/O
characters	86	Theme	shapes	18	Academic A/O
high	78	Reviews/Ratings	value	18	Academic A/O

appealing	77	Judgement	busy	17	Visuals
love	75	Judgement	application	17	Uninformative
theme	74	Theme	offers	17	Uninformative
enjoy	73	Judgement	received	17	Reviews/Ratings
worth	73	Cost/Benefit	check	17	More Information
learn	71	Academic A/O	childish	17	Judgement
terrible	69	Judgement	poorly	17	Judgement
educational	67	Academic A/O	hands	17	Buzzwords
positive	65	Judgement	answers	17	Benchmarks
images	64	Visuals	breaks	17	Benchmarks
skills	62	Academic A/O	evidence	17	Benchmarks
age	58	Suitability	education	17	Academic A/O
useful	57	Judgement	focus	17	Academic A/O
school	56	Academic A/O	teacher	17	Academic A/O
content	55	Uninformative	shown	16	Visuals
star	55	Reviews/Ratings	means	16	Uninformative
boring	55	Judgement	real	16	Uninformative
feel	54	Uninformative	lower	16	Suitability
information	54	More Information	reviewed	16	Reviews/Ratings
negative	54	Judgement	complicated	16	Judgement
classroom	53	Academic A/O	activity	16	Academic A/O
bit	52	Uninformative	helps	16	Academic A/O
younger	52	Suitability	tool	16	Academic A/O
lot	51	Uninformative	aspect	15	Uninformative
simple	51	Judgement	feature	15	Uninformative
games	51	Gamification	mention	15	Uninformative
interactive	51	Buzzwords	test	15	More Information
downloading	49	Uninformative	effective	15	Judgement
young	49	Suitability	unclear	15	Judgement
engage	49	Buzzwords	answer	15	Benchmarks
sounds	48	Uninformative	wrong	15	Benchmarks
concepts	48	Academic A/O	addition	15	Academic A/O
activities	47	Uninformative	animations	15	Academic A/O
appeal	46	Judgement	covers	15	Academic A/O
developed	46	Benchmarks	understand	15	Academic A/O

review	45	Reviews/Ratings	art	14	Visuals
comments	44	Reviews/Ratings	style	14	Visuals
waste	43	Cost/Benefit	actual	14	Uninformative
teachers	42	Academic A/O	experience	14	Uninformative
pay	41	Cost/Benefit	state	14	Uninformative
pretty	40	Uninformative	tasks	14	Uninformative
users	40	Reviews/Ratings	dogs	14	Theme
research	40	Benchmarks	majority	14	Reviews/Ratings
class	38	Academic A/O	fine	14	Judgement
interface	37	Visuals	scary	14	Judgement
fact	37	Uninformative	worst	14	Judgement
money	37	Cost/Benefit	created	14	Benchmarks
personalized	37	Buzzwords	development	14	Benchmarks
teaching	37	Academic A/O	bother	13	Uninformative
features	36	Uninformative	chance	13	Uninformative
rated	35	Reviews/Ratings	choose	13	Uninformative
play	35	Gamification	giving	13	Uninformative
child	35	Academic A/O	looked	13	Uninformative
colorful	34	Visuals	option	13	Uninformative
appears	34	Uninformative	prefer	13	Uninformative
learners	34	Academic A/O	product	13	Uninformative
design	33	Visuals	monsters	13	Theme
zombie	33	Theme	challenging	13	Suitability
awful	33	Judgement	difficult	13	Suitability
practice	33	Academic A/O	user	13	Reviews/Ratings
space	32	Theme	attractive	13	Judgement
score	32	Reviews/Ratings	avoid	13	Judgement
best	32	Judgement	unsure	13	Judgement
nice	32	Judgement	multimedia	13	Buzzwords
engaged	32	Buzzwords	covered	13	Academic A/O
aligned	32	Benchmarks	multiple	13	Academic A/O
attention	32	Academic A/O	topics	13	Academic A/O
sound	31	Uninformative	choice	12	Uninformative
suitable	31	Suitability	decide	12	Uninformative
educators	31	Academic A/O	options	12	Uninformative

pictures	30	Visuals	terms	12	Uninformative
appropriate	30	Suitability	higher	12	Reviews/Ratings
level	30	Suitability	popular	12	Reviews/Ratings
decent	30	Judgement	generic	12	Judgement
quality	30	Judgement	inviting	12	Judgement
experts	30	Benchmarks	spend	12	Cost/Benefit
teach	30	Academic A/O	eye	11	Visuals
character	29	Theme	approach	11	Uninformative
zombies	29	Theme	details	11	Uninformative
cute	29	Judgement	due	11	Uninformative
excellent	29	Judgement	provide	11	Uninformative
support	29	Academic A/O	questions	11	Uninformative
bright	28	Visuals	slightly	11	Uninformative
screenshots	28	Visuals	trust	11	Uninformative
idea	28	Uninformative	complex	11	Suitability
lots	28	Uninformative	grade	11	Suitability
right	28	Uninformative	reception	11	Suitability
highly	28	Reviews/Ratings	middle	11	Judgement
similar	28	Judgement	detailed	11	Benchmarks
standards	28	Benchmarks	meet	11	Benchmarks
student	28	Academic A/O	bored	11	Academic A/O
mediocre	27	Judgement	encourage	11	Academic A/O
makes	26	Uninformative	write	10	Written Information
people	26	Uninformative	claims	10	Uninformative
variety	26	Uninformative	effects	10	Uninformative
mixed	26	Reviews/Ratings	job	10	Uninformative
colors	25	Visuals	journey	10	Uninformative
visual	24	Visuals	mind	10	Uninformative
easy	24	Judgement	purchase	10	Uninformative
exciting	24	Judgement	stand	10	Uninformative
special	24	Judgement	monster	10	Theme
reason	23	Uninformative	levels	10	Suitability
stars	23	Reviews/Ratings	left	10	Reviews/Ratings
trial	23	More Information	dull	10	Judgement
promising	23	Judgement	simplistic	10	Judgement

blurb	22	Written Information	wasting	10	Cost/Benefit
screen	22	Visuals	aligns	10	Benchmarks
visually	22	Visuals	correct	10	Benchmarks
range	22	Uninformative	entertaining	10	Academic A/O
explore	22	More Information	hold	10	Academic A/O
trying	22	More Information	mathematical	10	Academic A/O
corrective	22	Benchmarks			

Note: Academic A/O refers to Academic Application/Outcome