WORKING PAPERS

Leveling Up to Literacy: Measuring the Impact of FTC's Online Advertising Literacy Game "Admongo"

David Givens

WORKING PAPER NO. 346

August 2020

BUREAU OF ECONOMICS
FEDERAL TRADE COMMISSION
WASHINGTON, DC 20580

Leveling Up to Literacy

Measuring the Impact of FTC's Online Advertising Literacy Game "Admongo"

David Givens, Bureau of Economics, FTC

8/14/2020

The author wishes to thank Jennifer Leach, Jan Pappalardo, Doug Smith, Ryne Marksteiner, Devesh Raval, Marc Luppino, Brian Rowe, Yan Lau and Dan Hosken at the FTC for helpful comments and suggestions. The analyses and conclusions set forth in this paper are those of the authors and do not necessarily reflect the views of other members of the Bureau of Economics, Bureau of Consumer Protection, other Commission staff, or the Commission itself.

Leveling Up to Literacy: Measuring the Impact of FTC's Online Advertising Literacy Game "Admongo"

David Givens, Economist, Division of Consumer Protection, Bureau of Economics, FTC

Abstract

This paper assesses the impact of an educational program designed by the Federal Trade Commission to teach advertising literacy to children ages 8-12. In a randomized experiment, children who played the program's instructional game scored significantly higher on an ad literacy test than students who did not play the game - approximately 9 percentage points higher on a 25-question test. The treatment group's advantage was evident on every question of the exam and remained virtually unchanged after controlling for demographic imbalances across groups. Features of the sample and the exam constrain the generalizability of results, however. In addition, a higher rate of attrition among those assigned to the treatment group introduces the possibility of unmeasured differences confounding our treatment-effect estimate. We discuss the implications of each challenge and offer a more-conservative estimate of Admongo's impact using an intent-to-treat approach. Despite these limitations, our study is one of only a small number to quantify the impact of an ad-literacy intervention using a randomized experiment.

I. Introduction

Children in the United States account for substantial purchasing power. Estimates vary, but studies by market research firms suggest that children aged 8-12 ("tweens") account for tens of billions of dollars in consumer spending annually, and that additional spending by parents and family members on this population ranges in the hundreds of billions of dollars per year (Mesa, 2005; Fromm, 2018; Chester, 2012). Approximately one-third of parents report that their children influence decisions about purchases for the household (National Retail Federation, 2019).

U.S. children also encounter a significant volume of advertising that has the potential to influence their preferences and purchase decisions. A 2007 report by the U.S. Federal Trade Commission's (FTC's) Bureau of Economics estimated that, each year, the average American child between the ages of two and 11 saw 25,600 television commercials (FTC, 2007). About 53% of tweens and 29% of teens reported that they try to buy products they have seen on TV or in the movies (Harris Interactive, 2006).

A policy issue arises because research shows that children are less capable than adults of understanding that the persuasive intent behind advertising influences the way products are represented (Livingstone and Helsper, 2006; Hobbs, 2004; Oates et al. 2003; Lapierre et al. 2017). Children may process advertisements without appreciating that the characterization of the product or service is entirely from the seller's point of view, which may differ from their own (Kunkel et al., 2004; Martin, 1997). Furthermore, due to the increasing presence of marketing embedded within movies, websites, television shows, video games and mobile apps, children may have increasing trouble differentiating ads from content (Kunkel et al., 2004; An et al. 2014).

An ability to identify ads, an awareness of their intent, and a familiarity with their persuasive techniques is referred to broadly as “advertising literacy” or “ad literacy.” In this paper, we examine an intervention called Admongo, launched in 2010 by the FTC to increase children’s advertising literacy. Drawing on the results of a randomized trial of Admongo carried out in 2016, we estimate the intervention’s impact and explore variation in ad literacy across demographic groups. This paper, to our knowledge, is one of only a handful to estimate quantitatively the effect of an ad literacy intervention using a randomized experiment.

The intervention evaluated in this paper is an educational game played online. The game was one component of a larger outreach program, which also included lesson plans, training videos and worksheets. An advertising literacy campaign like Admongo was a first for the FTC, but it was not the Commission’s first effort to mitigate the perceived harmful effects of marketing on children. For instance, in 1978, the FTC tried, unsuccessfully, to restrict television advertising to children (the so-called “KidVid” rules). That effort was motivated, in part, by concern that young children could not perceive the selling intent behind ads they encountered. The Commission has led child-related enforcement actions targeting deceptive toy advertisements, unauthorized charges in mobile apps, and privacy violations in online advertising (Beales, 2004).[1] The FTC has authored rules aimed at shielding children from what it deemed unfair intrusions by marketers.[2] Finally, FTC staff have carried out industry studies analyzing the marketing of products with potentially serious health effects for children.[3] In summary, over recent decades, the FTC has employed all of its consumer protection tools - enforcement, rulemaking, education and advocacy – in efforts to mitigate perceived harm to children from certain forms of marketing. Admongo, while a first in one sense, was also a continuation of an existing policy agenda.

Regulatory and financial barriers to running a randomized classroom study of the entire curriculum led us to narrow our evaluation to just the game, which could be studied more simply and cheaply, using an online panel of subjects. The game is a stand-alone teaching tool in which players advance to higher levels by correctly answering questions on advertising topics introduced via text, images and audio. In the evaluation of the game, a sample of 8-12 year-old children played to a required endpoint and then took an ad literacy test. A control sample of otherwise similar children took the same test but without exposure to the game. A market research firm recruited the sample, managed randomization and collected the results.

To preview results, we find that children who played the game, on average, scored approximately nine percentage points higher on a 25-question ad literacy test, compared to control students who did not play the game. Treatment students showed improved odds of answering correctly on every question of the test. The estimated differential varies only slightly if we control for demographic differences across groups. Under the assumption that the

1 For deceptive toy advertising, see, e.g.: Hasbro, Inc., (1993) or Mattel, Inc. (1971). For unauthorized charges, see, e.g. Apple, Inc. (2014) or Google, Inc. (2014). For online privacy violations (COPPA violations) see Google/YouTube (2019).

2 For example, the 900 Number Rule (1992) and Children’s Online Privacy Protection (“COPPA”) Rule (1999).

3 Industry studies include alcohol (1999, 2003, 2008, 2014); food and beverages (2008, 2012); and, violent movies, music and video games (2000 – 2009; seven studies). These industry studies documented existing practices and recommended improvements to self-regulatory efforts.

observable controls capture all outcome-relevant differences between control and treatment groups, these estimates represent an average treatment effect (ATE) for the population sampled.

These results have two important limitations, however. First, while the experimental setup allows for a precise treatment-effect estimate in the given context, it does not allow for generalizing to the broader population of U.S. 8-12 year olds or for predicting long-term changes in knowledge, attitudes or behavior. A lengthier intervention using a sample drawn from the general public and followed up with a prolonged evaluation would be required. Second, high attrition among participants assigned to the treatment group could mean that unobservable differences are confounding these treatment-effect estimates. To address this possibility, we present a quasi-intent-to-treat (ITT) estimator that assumes "treatment-quitters" would have derived zero benefit had they played the game as directed.

The paper proceeds as follows. In Section II, we offer background on children's advertising literacy by examining conceptualizations offered by the marketing and psychology literatures in recent decades. In Section III, we present the experiment design, including a brief description of the Admongo game itself. In Section IV, we present analysis of the study results. Finally, in Section V, we offer further discussion along with our concluding thoughts.

II. Children's Advertising Literacy – Concepts from the Literature

A broad marketing and psychology literature on childhood ad literacy has yielded a few consistent themes but also many findings that vary with the way ad literacy is defined and measured. Indeed, defining ad literacy has been an issue that has dogged the literature for decades. In this paper, we do not advocate for one definition over another. We offer the following brief literature review simply to provide foundational concepts of advertising literacy so that the content of FTC's program can be understood better.

At a minimum, the concept of ad literacy includes the ability to identify a commercial message – for example, the ability to distinguish ads from programming, while watching television. On this score, Levin et al. (1982), for example, find that preschoolers as young as three years old reliably distinguish television ads from programming.

Most research equates ad literacy with something deeper than mere recognition of advertisements – i.e., some level of understanding of the advertiser's intentions. Donohue et al. (1980) found that children as young as three to six years old understood that an animated television ad was urging them to buy cereal. Martin (1997) and Andronikidis and Lambrianidou (2010) looked separately at children's understanding of the informative versus the persuasive intents of advertisers. Andronikidis and Lambrianidou found that children could perceive advertising's informative role at an earlier age than they could understand its persuasive intent. Martin highlighted the same qualitative result across multiple studies in her meta-analysis. In a related vein, Pine and Veasey (2003) examine children's understanding of the rhetorical and tactical role of promotional messages more broadly. They find that children as young as four years old understand the appropriateness of promotional messages, in both advertising and non-advertising contexts. Their study is interesting for its evidence on young children's awareness of the connection between intentionally biasing a message and achieving a desired effect.

Still other work distinguishes between literacy in the sense of conceptual knowledge and literacy in the sense of a critical attitude. Opree and Rozendaal (2015) find that it is children's "attitudinal" advertising literacy (their feelings, opinions, critical posture), not their "conceptual" ad literacy (their understanding of persuasive intent, etc) which is correlated with children's ability to resist advertising appeals.

Opree and Rozendaal (2015) highlight another focus of research, which is the extent to which one's knowledge about advertising mediates one's responses to advertising. Because research into children and advertising stems largely from a concern that children lack the defenses necessary to resist advertisers' appeals, the role of literacy as a "defense mechanism" is a central concern. In their study of children's responses to advergames, Reijmersdal et al. (2012) find that "…persuasion knowledge (i.e. knowledge of the commercial source of the game and its persuasive intent) did not influence cognitive or affective responses to the brand or game." In other words, just teaching children about persuasive intent did not make them less susceptible to advertisers' appeals.

Rozendaal et al. (2016) suggest that putting children on guard against being manipulated by ads could be more effective; they found that forewarning children of an upcoming ad's "manipulative" intent reduced "product desire" following the ad, but that forewarning them only of the ad's "commercial" intent did not. The authors surmise that the difference has to do with children's emotional response to the anticipation of being manipulated. In related work, Rozendaal et al. (2011) propose a three-part ad literacy framework of concepts, attitudes and performance, arguing that merely equipping children with concepts doesn't make them capable of "…using that knowledge as a defense".

Another persistent question in the literature is the age at which advertising literacy awakens in children. Different studies have found different age gradients, again depending on how literacy is defined and measured. One regularity is that the correlation between age and advertising literacy tends to be stronger in studies using verbal measures because very young children have more trouble expressing themselves verbally. In other words, studies that utilize verbal measures may understate advertising literacy in very young children. Pine and Veasey (2003), Donohoe et al. (1980) and Martin's (1997) meta-analysis all report variants of this result.

While some advertising literacy research has focused on building children's resistance to marketing messages, the Admongo campaign did not have this aim. Rather, it sought to teach interpretive skills to a population (children) who were being targeted in new ways and who possessed the least experience in thinking critically about advertising. Abundant economic research has demonstrated that truthful advertising can perform helpful roles for consumers.[4] The

[4] Stigler's (1961) seminal article on the economics of information highlights the direct informational benefit of advertising in lowering search costs by making consumers aware of the existence of sellers and the prices they charge. Nelsen (1974) highlights the ability of advertising - even ads lacking specific information on price or quality - to signal quality to consumers. Ippolito and Mathios (1989, 1990, 1995) present empirical evidence for advertising's ability to disseminate novel health information and move consumers towards healthier diets. Finally, empirical work by Benham (1972), and Milyo and Waldfogal (1999), and theoretical results in Bagwell and Lee (2010) demonstrate that advertising, despite representing an added expense to firms, can lead to lower prices and higher consumer surplus than regimes where advertising is restricted. Finally, by being positioned adjacent to

objective of Admongo was to help children understand advertising's techniques and objectives, and extract from ads the information that would enable them to make smarter purchasing decisions.

III. Study Design

In the spring and summer of 2016, the FTC conducted a study of the Admongo game's effectiveness in boosting ad literacy among children 8-12. We begin our discussion of that study with a brief description of the game itself.

A. The Admongo Game

The Admongo educational game was designed for use by elementary and middle school students and introduced players to topics such as points of view, ad targeting, persuasive techniques, and implicit and explicit claims.[5]

Figure 1. The Admongo game begins by having players create an avatar.

The game's substantive content begins by having players identify ads and articulate what ads are trying to get audiences to do. As the game progresses, players are asked to decode and deconstruct ads in order to see how they are made appealing for different target markets. A recurring message in the game is Admongo's three-question heuristic for evaluating ads: 1) "Who is responsible for the ad?" 2) "What is the ad actually saying?" and 3) "What does the ad want me to do?"

popular content, advertising can reach more consumers than government outreach or the scientific press, making ads a powerful channel for new health information (Ippolito and Mathios; 1989, 1990, 1995; Calfee and Pappalardo, 1991).

5 The FTC took the *Admongo* game off the Internet in June 2019. The game had been built on Adobe Flash, software that is being phased out in 2020. The game's content and lesson plans can still be found at: www.consumer.ftc.gov/admongo.

At the end of the game, players are asked to create their own advertisement for a particular target audience, drawing on what they have learned in the game. Thus, a central part of the educational strategy in the game seems to have been to show children the perspective of an advertiser, walking them through the process of targeting and producing an ad, so that children thereby will come to understand better the origin and meaning of advertisements that they themselves encounter.

Figure 2. Early levels of the Admongo game highlight the presence of advertisements in kids' surroundings, such as on the sides of municipal buses.

Figure 3. The game highlights ways in which advertisements can sometimes be difficult to distinguish from content.

The game, launched in 2010, could be played through a web browser on a Mac or Windows-based computer but not on a mobile device.

Note that study participants who played the game did not necessarily complete the game or encounter all of its educational content. To mitigate attrition, the FTC asked only that participants complete level 2.1, though players could continue further if desired.

B. Sampling

To carry out the evaluation, the FTC contracted with a market research firm to recruit an online sample of 8-12 year-old children. Recruits were randomized to control or treatment status and led through their respective regimens: game-play + test completion for the treatment group; test completion (only) for the control group. For each participant, overall performance on the test (i.e., overall % correct) as well as performance on each question were recorded. Basic demographic information on each participant and her household also were captured.

i. Overview

Study participants were recruited from the contractor's pre-existing panel of participating households, which had previously responded to solicitations to participate in marketing studies in return for compensation. The target number of participants for this evaluation was 800 children, evenly divided between control and treatment. Ultimately, a sample of n=791 (401 control and 390 treatment) followed protocol and completed the ad literacy test.

The target sample size of 800 was chosen as a trade-off between cost and statistical power. Before settling on a sample size, a power analysis was conducted to determine the minimum detectable effect (MDE) of the experimental design, i.e. – the smallest difference in mean test score across control and treatment groups for which there is at least an 80% probability of rejecting the null hypothesis of no effect at the 5% significance level. For a comparison of the entire treatment group and entire control group, with n=400 in each group, a conservative estimate of the MDE is 4 points out of 100, or approximately one question on the FTC's 25-question ad literacy test.[6] For a comparison of control and treatment means by sex or by age category (n=200 in each age-treatment or sex-treatment status group), a conservative estimate of the MDE is 5.6 points, or approximately 1.5 questions.[7] Finally, for a comparison of control and treatment means by age-sex category (n=100 in each age-sex-treatment status group), a conservative estimate of the MDE is 8 points, or approximately two questions on a 25-question test.

Parents or caregivers who agreed to participate were directed first to fill out a demographic questionnaire (a "screener"). If the household met eligibility requirements (at least one child 8-12, a computer with internet access, etc.), the household was randomized to either the control or treatment arm of the experiment. Sampling was stratified to approximate the U.S. 8-12 year-old population along the dimensions of sex, age, household income, race/ethnicity and Census region.

[6] The estimate is based on the assumption of a control group population with a mean score of 70% correct and a standard deviation of 20 percentage points.
[7] Subjects were divided into two age categories: 8-10 years, and 11-12 years.

ii. Representativeness

The n=791 sample that completed protocol (what we'll call the "completer" sample) is unlikely to constitute a nationally representative sample of Admongo's target population for three reasons. First, the sample was drawn from a panel of paid research subjects instead of the public at large. Second, panel households could refuse to participate in any given study, including this one. And, third, participants could abandon a study at any point prior to completion.

These factors suggest that the distribution of unobserved, outcome-relevant characteristics among sample members may differ from those distributions among the broader 8-12 year-old U.S. population. For example, subjects' previous experience in market research studies may heighten their baseline awareness of advertising and thereby bias downward our estimate of the treatment effect of playing Admongo. Because the overall representativeness of our sample is ultimately unknowable, we do not generalize effect-size estimates derived in this paper to the broader audience of all 8-12 year-old Americans.

Table 1, below, compares the distribution of select variables in our Admongo completer sample with the corresponding distributions in the wider U.S. population.

Admongo Sample vs. U.S. Population

	(1) Admongo sample (completers, n=791)	*Comparison Population Data from U.S. Census* (2) U.S. 8-12 yr. old population[1]	(3) U.S. households with one or more children[2]	(4) U.S. adults, 25 yrs and older[3]	(5) U.S. householders, 25 yrs and older[2]
Child's Race/Ethnicity					
American Indian/AK Native	0.6%	1.6%			
Asian	4.4%	5.2%			
Black	12.8%	14.9%			
Native HI/Pacific Islndr.	0.4%	0.3%			
White	76.0%	72.8%			
Other	4.4%	5.2%			
Annual Household Income					
<$40K	24.2%		28.3%		
$40K-$50K	10.9%		7.5%		
$50K-$75K	23.0%		15.7%		
$75K-$100K	16.3%		15.7%		
>$100K	25.7%		32.9%		
Parent/Caregiver's Educational Attainment					
<High school grad.	0.4%			13.3%	9.4%
High school grad. (incl. GED)	10.9%			27.8%	26.4%
Some college, no degree	23.8%			21.1%	17.2%
2-yr. degree	13.0%			8.1%	11.0%
4-yr. degree	31.1%			18.5%	21.9%
Master's deg. or higher	20.9%			11.2%	21.9%

Sources:

[1]U.S. Census Bureau, Annual Estimates of the Resident Population by Sex, Single Year of Age, Race, and Hispanic Origin for the United States, 2016 population estimates. See <https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk>

[2]U.S. Census Bureau, Current Population Survey, 2017 Annual Social and Economic Supplement, Table FINC-01. Selected Characteristics of Families by Total Money Income in: 2016. See <https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-finc/finc-01.html>.

[3]U.S. Census Bureau, 2012-2016 American Community Survey 5-year estimates, Educational Attainment. See <https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk>

Table 1. Admongo sample vs. U.S. population, select variables.

The Admongo sample's race and ethnicity profile matches the U.S. 8-12 year-old profile fairly closely. This is not surprising, given the stratification of the sampling design. With respect to household income, the Admongo sample has relatively fewer households at either income extreme, compared to the national distribution of households with one or more children.

Adults in Admongo households tend to be more highly educated than typical U.S. adults, although there is some uncertainty over the ideal comparison population. Column 4 presents educational attainment for all U.S. adults age 25 and older, while column 5 presents the same concept for all U.S. *householders* age 25 and older. Since an Admongo parent/caregiver resides in a household with children but is not necessarily a householder (i.e., the person in whose name the residence is owned or rented), we cannot be certain which figures represent the appropriate comparison group. We show educational attainment data from both populations to bracket the potential scenarios. Under either comparison, the responsible adults in Admongo-household appear to be more likely to have finished high school and to have attained a four-year college degree than the comparable U.S. population.

iii. Threats to randomization

Because the sampling plan called for identically stratified control and treatment samples, assignment of any particular participant was, in most cases, not a coin-flip. That is, an eligible respondent with a given demographic profile was assigned on a least-filled basis - i.e., to the treatment arm that was, at that moment, farthest from its sampling quota for subjects like her. In this way, the probability of assignment to treatment varied across participants and over time.

As a general rule, the probability of assignment to the treatment group rose over time because attrition was substantially higher among participants assigned to treatment than it was among those assigned to control.[8] The differential attrition likely stemmed from the additional time commitment required of treatment subjects, although compensation from the contractor for the treatment group was also higher. Whatever the reason, one consequence was that the probability of being assigned to the treatment group grew over time because that group was perpetually in greater need of subjects. (Appendix Figure A2 presents a graphical depiction of the fielding process over time.)

The increase over time in assignment to treatment could have compromised the balance of control and treatment groups if later participants were systematically different from earlier participants in unmeasured ways. However, because fielding lasted only three months, we see little reason to suspect this phenomenon undermined comparability in ways we cannot measure.

Difficulty attracting low-income participants (annual household income < $40K) presented another challenge to the sampling plan. Recruitment and data collection began on March 28, 2016, but as of May 16, just 51% of the targeted number of low-income participants had completed protocol, vs. 93% of the targeted number of children from all other income strata, collectively. Because the treatment group, in general, suffered from higher attrition, the low-income stratum in that group was particularly far from its target; just 23% of the treatment group's low-income quota had been met as of May 16, compared to 79% of the control group's quota.

To expedite fielding, income quotas were dropped from the sampling plan on June 1 for both treatment and control groups. After continued slow uptake in the treatment group, all demographic quotas were dropped on June 15. The requisite threshold of 400 subjects was reached for the control group one day later (June 16) and for the treatment group on June 24.[9] Reaching the target sample size ultimately involved recruiting 4,230 households over the course of fielding. Attrition remained concentrated in the treatment group throughout, with 81% of individuals who were assigned to treatment dropping out before completing the ad literacy test. Among control-group assignees, the attrition rate was substantially lower, at 37%.

---

[8] We use "attrition" here to describe the situation in which a prospective participant who was recruited, found eligible, and randomized to control or treatment status, quit at any point before completing the ad literacy test.

[9] See the Appendix, Figure A2, for a graphical depiction of the recruitment process over time.

Covariate balance suffered in some measure from the asymmetric attrition and from the ensuing decision to abandon sampling quotas.[10] Table 2, below, shows that treatment completers tended to come from wealthier, more highly educated households, and were more likely to be white, have married parents, and have multiple siblings. Control completers were more likely to be black, older (i.e., age 11 or 12), an only child, have a single parent, have parents who were employed full-time, and live in the South. Control students also were more likely to be sixth graders, perhaps because more of that group had been recruited earlier in the spring (when a twelve-year-old was more likely to be in sixth grade, as opposed to fifth grade).

[10] See Appendix Table A2, showing demographic characteristics of control and treatment assignees who entered before and after income quotas were lifted on June 1.

Admongo Sample, "Completers"

	Control	Treatment	Difference	p-value
N	401	390		
HH income < $40K	30.8%	17.3%	13.5%	0.000
HH income > $50K	56.5%	73.8%	-17.3%	0.000
HH income > $100K	22.1%	29.4%	-7.3%	0.021
Census South	37%	32%	5.1%	0.129
Census Northeast	18%	18%	0.3%	0.926
Census Midwest	22%	26%	-4.7%	0.121
Census West	23%	24%	-0.7%	0.827
Child white	71.8%	80.3%	-8.4%	0.006
Child black	14.7%	10.8%	3.9%	0.097
Child Asian	5.2%	3.6%	1.6%	0.260
Child hispanic	14.7%	12.3%	2.4%	0.323
Child American Indian/AK Native	0.7%	0.5%	0.2%	0.676
Child HI Native/Pacific Islander	0.2%	0.5%	-0.3%	0.547
Child other race	7.2%	4.4%	2.9%	0.084
Num. siblings	1.3	1.5	-0.21	0.007
Only child	22%	15%	6.3%	0.023
2+ siblings	31%	39%	-7.3%	0.032
Child male	50%	51%	-0.6%	0.856
Child grade (mean)	4.67	4.54	0.13	0.127
Child grade 3	25%	29%	-4.3%	0.174
Child grade 4	17%	17%	-0.2%	0.932
Child grade 5	23%	23%	0.4%	0.904
Child grade 6	34%	30%	4.2%	0.212
Child age (mean)	10.3	10.2	0.10	0.294
Child age 8	16%	15%	0.8%	0.751
Child age 9	16%	20%	-3.8%	0.165
Child age 10	18%	19%	-1.5%	0.580
Child age 11	24%	23%	0.6%	0.838
Child age 12	26%	23%	3.9%	0.206
Child homeschooled	5%	9%	-4.0%	0.032
Child public school	86%	83%	2.7%	0.293
Child private school	9%	7%	1.3%	0.505
Weekend test taker	8%	16%	-7.2%	0.002
Night test taker	12%	4%	8.4%	0.000
Morning test taker	29%	26%	3.0%	0.345
Afternoon test taker	23%	36%	-12.9%	0.000
Evening test taker	36%	34%	1.6%	0.648
Parents married	70%	78%	-8.1%	0.009
Parent single	10%	6%	3.8%	0.049
Parent bachelor's +	49%	55%	-5.7%	0.106
Parent FT employed	62%	56%	6.2%	0.076
Parent PT employed	10%	15%	-4.9%	0.037
Parent unemployed	4%	2%	1.7%	0.157
Parent homemaker	21%	24%	-3.1%	0.287
Parent male	23%	24%	-1.2%	0.701
Parent master's +	21%	21%	0.2%	0.951
Parent divorced/separated	11%	9%	1.5%	0.484
Parent 40+	54%	51%	3.6%	0.311
Parent 50+	11%	8%	3.5%	0.090

Table 2. Covariate balance in control and treatment groups in the sample of completers.

C. The ad literacy test

The study's endpoint was the participant's score on an ad literacy test written collaboratively by FTC staff and outside marketing consultants. The test took students approximately 15-20 minutes to complete.

The ad literacy test consisted of thirteen true-false questions, twelve multiple-choice questions, 39 spot-the-ad questions and nine subjective questions about participants' attitudes and experiences. The full test can be found in the Appendix of the paper.

The 25 true-false and multiple-choice questions focus on recognizing various persuasive techniques; understanding ad targeting; identifying ads in their various forms; interpreting explicit and implied claims; understanding advertisers' responsibilities under the law; and understanding the financial relationships behind advertising (its role in supporting content, paid testimonials, etc.). These questions elicit a mix of specific factual knowledge (e.g., the meaning of "target audience" or the name of a specific persuasive technique) and higher-order critical thinking skills (e.g., understanding why an advertiser might use cross-promotion; predicting the environment most suited for reaching an ad's intended audience, etc.).

The multiple-choice section centered on interpretation of mock advertisements created especially for the test. The ads target roughly the 8 - 18 year-old age range and present products likely to garner the attention of 8 - 12 year-old children. One ad is for a yogurt drink; the second is for a video game system; the third is for a brand of casual shoes. The questions that accompany the mock ads ask students to evaluate the claims, persuasive techniques, targeting, and presence or absence of product information.

The nine opinion-based questions touch on the ways that children gather information about products they want, how children's experiences with products compare to the impressions given in advertisements, and whether children receive guidance from their parents about advertising.

The spot-the-ad section asks students to identify all the advertisements in a cartoon graphic. One concern with the results from this section is that false positives were not recorded. That is, we observe when a student fails to pick out something that was an ad but not when she falsely picks out something that wasn't an ad.

Because of the subjectivity in the opinion questions, and the technical complications with the spot-the-ad questions, we treat the 25 true-false and multiple-choice questions as the main outcome of interest in the analysis that follows.

D. Procedure

Treatment students were instructed to play the Admongo online game from their homes, to reach a certain level within the game (Level 2.1), and then to complete the advertising literacy test. Control students were provided a link directly to the test and told to complete it, and they were not told about Admongo or its connection to the test. To ensure the integrity of the evaluation, the market research contractor screened out any panelist who reported knowledge about Admongo prior to this study.

IV. Results.

The sampling plan called for n=400 participants in control and in treatment. In the end, complete data was collected for 391 treatment participants and 401 control participants.

A. Unadjusted Results

We begin our analysis by looking at raw differences in test scores between control and treatment groups, without adjusting for covariates. Treatment students scored significantly better than control students on the test as a whole, and on all test subsections, as Table 3, below, indicates.

Unadjusted Score Differences, Control vs. Treatment

	control	treatment	difference	p-value
N	401	390		
All questions	0.452	0.591	0.140	0.000
Std. Err.	0.008	0.008		
T/F & M/C only	0.693	0.79	0.098	0.000
Std. Err.	0.007	0.006		
Spot-the-ad only	0.297	0.464	0.167	0.000
Std. Err.	0.011	0.012		

Table 3. Mean scores on the ad literacy test: entire test, and by sub-section.

Histograms presented in Figures 5-7, below, give a fuller picture of the test score distribution in control vs. treatment.

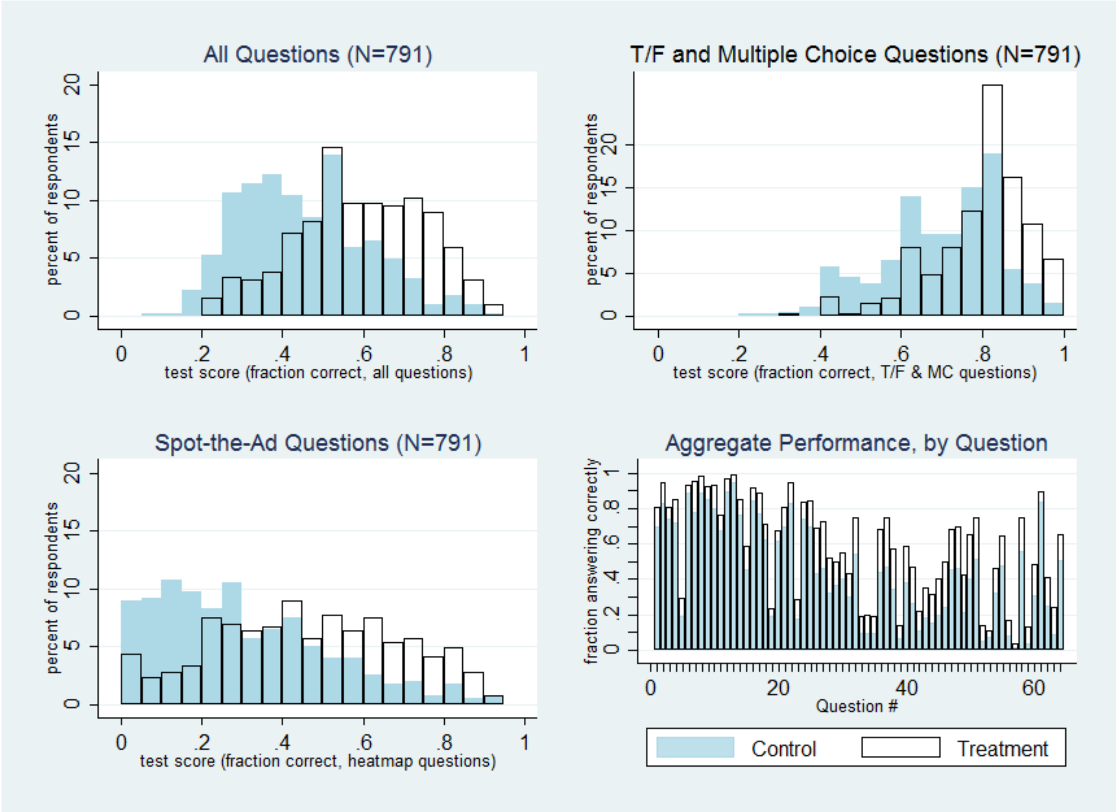


Figure 4. Test score distributions, control vs. treatment.

The plot in the lower right of Figure 4 suggests that the effect of playing Admongo is comprehensive, across the entire range of questions on the test. That is, exposure to Admongo appears to make a child more likely to answer correctly on every individual question. Again – the test is included for reference in the Appendix of this paper.

Looking at specific test sections, treatment appears to have provided more of an advantage with spot-the-ad questions than with the T/F and multiple-choice questions. Treatment children averaged 16.7 percentage points higher than control children in the spot-the-ad section, but only 9.8 points higher than control children on the T/F and multiple-choice sections (p-value = 0.00 in both comparisons). These results suggest that one of the strongest impacts of playing the Admongo game could be a heightened awareness of the forms that advertising can take in one's surroundings. On the other hand, we cannot rule out the possibility that the treatment group's advantage on this section stems primarily from its exposure to some of the very same product images in the game's tutorials. Partly for this reason, we do not include spot-the-ad results in the analysis in coming sections.

The next two figures focus specifically on the T/F and multiple-choice sections, as these constitute the basis of analysis in the rest of the paper. Figures 5 and 6 look at the extent to which the treatment affected varied across sub-populations in the data. Figure 5 suggests an effect is present among whites, non-whites, lower and middle/upper-income, male and female. Figure 6 breaks down participants by a different metric: the time of day and day of the week (weekday vs. weekend) on which they took the test. There is evidence of a potential treatment effect at all times of day and on both weekdays and weekends.

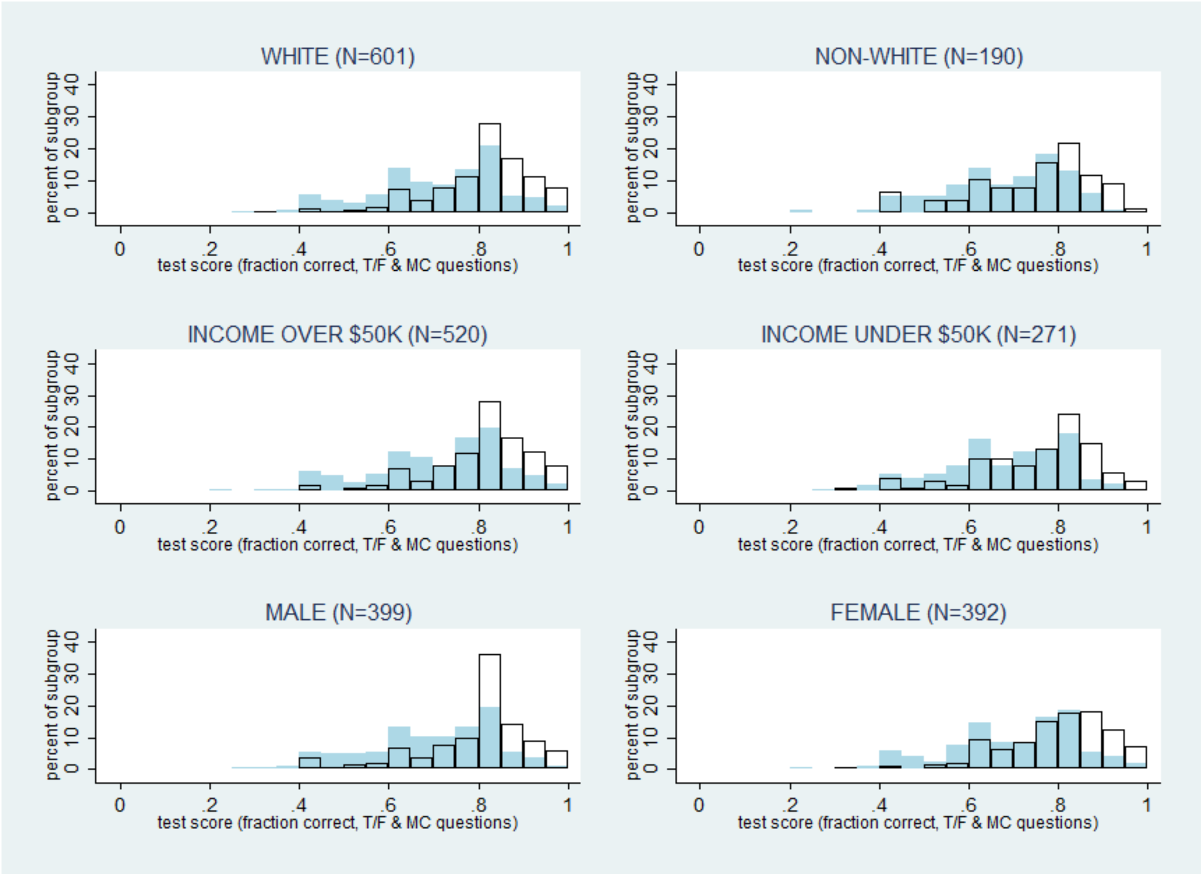


Figure 5. Test score distributions, control vs. treatment, by covariate groups.

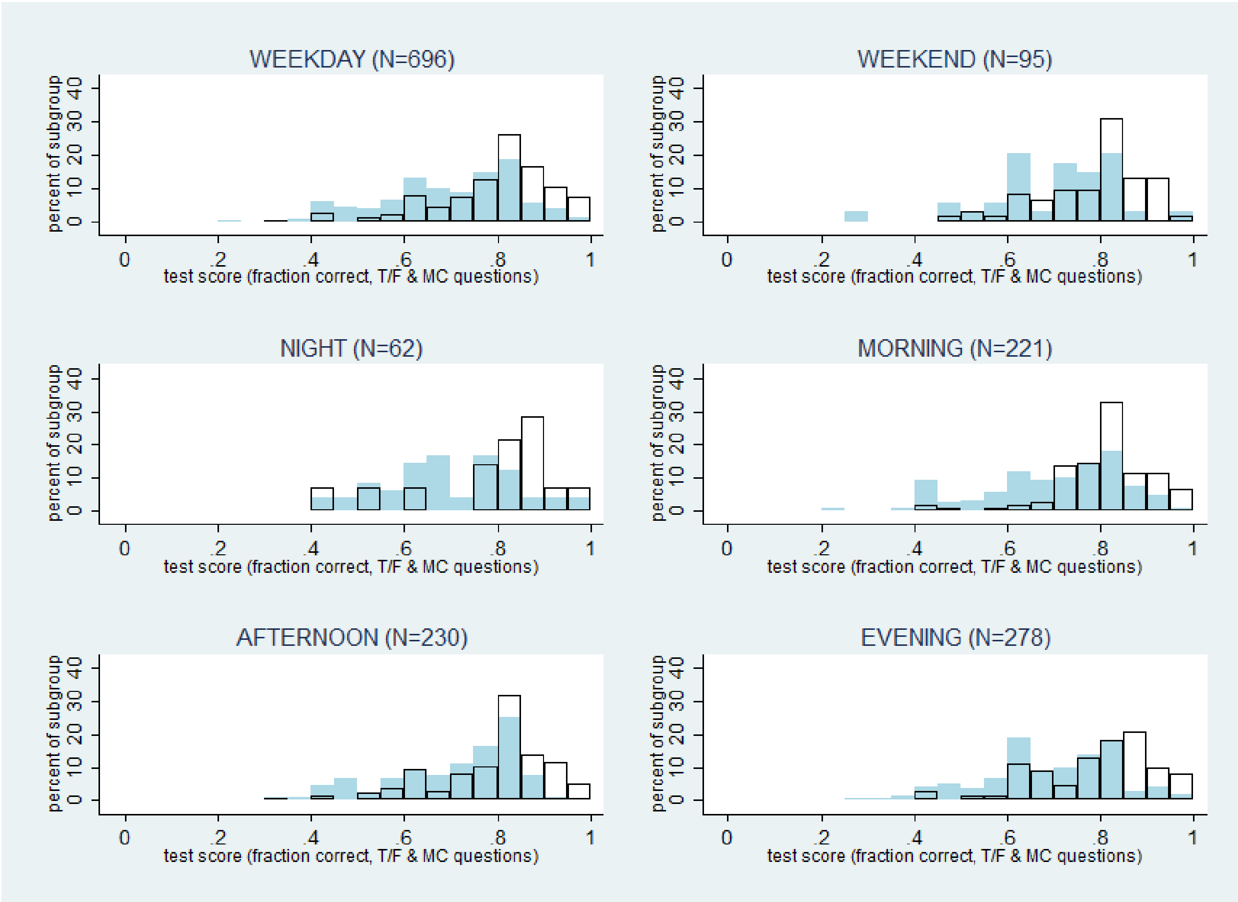


Figure 6. Test score distributions, control vs. treatment, by test time and day ("night" = midnight – 6:00 am; "morning"=6:01 am – noon, etc.)

While the histograms suggest that playing Admongo had a positive effect on test performance across a variety of sub-populations, the plots also suggest that the distribution of ad literacy at baseline may differ across groups, and that the effect of Admongo also could differ across groups. And, because they hold constant only one covariate at a time, the histograms do not adjust for other within-group differences between control and treatment. We explore the potential role of covariates more formally through regressions presented in the next section.

B. Adjusted Results

In this section, we assess the treatment effect of Admongo more rigorously by controlling for potentially confounding variables. Demographic controls are necessary because, despite random assignment, our preliminary look at the sample indicated that control and treatment groups were balanced on some, but not all the demographic variables. Due to differential attrition, the lifting of sampling quotas, and random sampling variation, the treatment group had a greater concentration of upper-income children, younger children (i.e., 3rd and 4th graders), children from married households, white children, and children with two or more siblings. Even if sampling had proceeded as planned and all demographic quotas had been adhered to, it is unlikely the joint distribution of demographic variables would have been identical in both control and treatment samples. Regression analysis enables us to control for the confounding differences across experimental arms and estimate specifically the impact of the Admongo intervention on ad literacy. The regressions also allow us to explore differences in baseline ad literacy and in the treatment effect across demographic groups.

We begin with Table 4, below, which presents coefficient estimates from a set of OLS regressions of the test score on a treatment dummy and a variety of demographic controls. In Table 4 and all analysis that follows, the dependent variable is the combined score (percentage correct) on the 25 true/false and multiple-choice questions.

Dependent variable = % correct out of 25 T/F, M/C questions	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treated	0.098***	0.098***	0.098***	0.094***	0.092***	0.096***	0.092***	0.089***	0.092***	0.090***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Male		-0.005								
		(0.01)								
Grade 4			0.003				0.004	0.009	0.003	0.009
			(0.01)				(0.01)	(0.01)	(0.01)	(0.01)
Grade 5			0.008				0.007	0.009	0.006	0.008
			(0.01)				(0.01)	(0.01)	(0.01)	(0.01)
Grade 6			0.009				0.006	0.007	0.007	0.008
			(0.01)				(0.01)	(0.01)	(0.01)	(0.01)
Amer. Ind./AK Ntv.				-0.062				-0.056		-0.053
				(0.06)				(0.06)		(0.06)
Asian				-0.086***				-0.092***		-0.093***
				(0.02)				(0.02)		(0.02)
Black				-0.031*				-0.020		-0.022
				(0.01)				(0.01)		(0.01)
Ntv. HI/Pac. Isldr.				-0.060				0.014		0.008
				(0.08)				(0.09)		(0.09)
Other				-0.027				-0.012		-0.016
				(0.02)				(0.02)		(0.02)
HH income: $40K-$50K					0.006		0.006	0.004	0.004	0.001
					(0.02)		(0.02)	(0.02)	(0.02)	(0.02)
HH income: $50K-$75K					0.036**		0.037**	0.037**	0.034*	0.033*
					(0.01)		(0.01)	(0.01)	(0.01)	(0.01)
HH income: $75K-$100K					0.039*		0.039*	0.040*	0.033*	0.032*
					(0.02)		(0.02)	(0.02)	(0.02)	(0.02)
HH income: Over $100K					0.054***		0.054***	0.053***	0.044**	0.041**
					(0.01)		(0.01)	(0.01)	(0.02)	(0.02)
Guardian: Some college / trade school						0.019			0.014	0.016
						(0.02)			(0.02)	(0.02)
Guardian: 2-yr degree						-0.006			-0.014	-0.009
						(0.02)			(0.02)	(0.02)
Guardian: 4-yr degree						0.028			0.012	0.015
						(0.02)			(0.02)	(0.02)
Guardian: Master's or higher						0.045*			0.024	0.031
						(0.02)			(0.02)	(0.02)
constant	0.693***	0.695***	0.687***	0.704***	0.666***	0.671***	0.662***	0.669***	0.656***	0.662***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
R-squared	0.117	0.118	0.118	0.139	0.142	0.131	0.142	0.161	0.148	0.168
N	791	791	791	791	774	791	774	774	774	774

Standard errors in parentheses.

Table 4. OLS regressions of test score on controls and Admongo treatment dummy.

One overarching result from Table 4 is that the estimated treatment effect from Admongo is sizable and robust to controlling for the potential confounders that we can measure. Playing the Admongo game appears to raise a student's expected score on the 25-question test by about nine percentage points, or roughly the equivalent of two additional correct responses. Whether we control for a child's grade level, race, household income, parental education - or all of these - the estimated effect of playing Admongo remains relatively constant and statistically significant at the 1% level. In other specifications not presented here, we also have controlled for parental employment status, month of participation, whether participation occurred after income quotas were lifted, the timing of participation (morning, afternoon, etc.), weekend vs. weekday participation, the household's Census region, number of siblings, and school type (public, private, home-schooled). In all cases, coefficient estimates on the Admongo treatment dummy remain in the 8-10 percentage-point range and significant at the 1% level.

Another basic result from Table 4 is the very limited ability of demographic controls to explain test-score variation. Virtually none of the student, guardian or household-level characteristics enter regressions significantly. To be sure, there are suggestive patterns: older children (grades 5-6), children from households with higher incomes, children of guardians who are college graduates and white children all appear to enjoy small score advantages. Children of Asian descent seem to fare relatively poorly. However, in almost all instances, the coefficient estimates suggesting these patterns are not statistically significantly different from zero. In those regressions excluded from Table 4 that were mentioned above, no explanatory variable entered significantly except "two or more siblings," which was associated with a negative effect of approximately two percentage points in some specifications. We note the linkage here but do not venture a guess as to the reason for the relationship.

It is curious that demographic variables can explain so little of the score variation. This result may be, in part, the by-product of drawing a sample from a marketing research panel rather than from the public at large, where variation in potential test performance could be greater. However, before dismissing possible relationships between demographic variables and the outcome, we explore the relationships more deeply using interactions between controls and the Admongo treatment dummy. These results are shown in Table 5, below.

Dependent variable = % correct out of 25 T/F, M/C questions	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treated=1	0.098***	0.127***	0.101***	0.086***	0.097***	0.117***	0.127***	0.117***	0.129***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Grade 4		0.016				0.014	0.021	0.013	0.021
		(0.02)				(0.02)	(0.02)	(0.02)	(0.02)
Grade 5		0.034				0.034	0.039*	0.032	0.038
		(0.02)				(0.02)	(0.02)	(0.02)	(0.02)
Grade 6		0.029				0.029	0.035*	0.030	0.037*
		(0.02)				(0.02)	(0.02)	(0.02)	(0.02)
Treated=1 # Grade 4		-0.022				-0.019	-0.021	-0.019	-0.020
		(0.03)				(0.03)	(0.03)	(0.03)	(0.03)
Treated=1 # Grade 5		-0.051				-0.052	-0.058*	-0.052	-0.058*
		(0.03)				(0.03)	(0.03)	(0.03)	(0.03)
Treated=1 # Grade 6		-0.040				-0.047	-0.057*	-0.048	-0.058*
		(0.02)				(0.03)	(0.03)	(0.03)	(0.03)
Amer. Ind./AK Ntv.			-0.061				-0.054		-0.052
			(0.08)				(0.08)		(0.08)
Asian			-0.068*				-0.075*		-0.077*
			(0.03)				(0.03)		(0.03)
Black			-0.012				0.003		0.002
			(0.02)				(0.02)		(0.02)
Ntv. HI/Pac. Isldr.			-0.181				0.016		0.018
			(0.13)				(0.09)		(0.10)
Other			-0.023				-0.018		-0.020
			(0.03)				(0.03)		(0.03)
Treated=1 # Amer. Ind./AK Ntv.			-0.001				-0.010		-0.014
			(0.12)				(0.12)		(0.12)
Treated=1 # Asian			-0.042				-0.047		-0.046
			(0.05)				(0.05)		(0.05)
Treated=1 # Black			-0.044				-0.052		-0.051
			(0.03)				(0.03)		(0.03)
Treated=1 # Ntv. HI/Pac. Isldr.			0.179				0.000		0.000
			(0.16)				(.)		(.)
Treated=1 # Other			-0.007				0.004		0.003
			(0.04)				(0.05)		(0.05)
HH income: $40K-$50K				0.021		0.020	0.017	0.017	0.014
				(0.02)		(0.02)	(0.02)	(0.02)	(0.02)
HH income: $50K-$75K				0.038*		0.041*	0.042*	0.037	0.038
				(0.02)		(0.02)	(0.02)	(0.02)	(0.02)
HH income: $75K-$100K				0.022		0.022	0.025	0.013	0.015
				(0.02)		(0.02)	(0.02)	(0.02)	(0.02)
HH income: Over $100K				0.046*		0.047*	0.046*	0.032	0.030
				(0.02)		(0.02)	(0.02)	(0.02)	(0.02)
Treated=1 # HH income: $40K-$50K				-0.035		-0.034	-0.027	-0.032	-0.025
				(0.04)		(0.04)	(0.04)	(0.04)	(0.04)
Treated=1 # HH income: $50K-$75K				-0.001		-0.004	-0.004	-0.002	-0.003
				(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
Treated=1 # HH income: $75K-$100K				0.033		0.035	0.031	0.043	0.038
				(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
Treated=1 # HH income: Over $100K				0.015		0.017	0.018	0.027	0.027
				(0.03)		(0.03)	(0.03)	(0.03)	(0.03)
Guardian: Some college / trade school					0.014			0.007	0.007
					(0.02)			(0.02)	(0.02)
Guardian: 2-yr degree					0.001			-0.010	-0.006
					(0.03)			(0.03)	(0.03)
Guardian: 4-yr degree					0.030			0.017	0.020
					(0.02)			(0.02)	(0.02)
Guardian: Master's or higher					0.046			0.032	0.037
					(0.02)			(0.03)	(0.03)
Treated=1 # Guardian: Some college / trade school					0.011			0.017	0.013
					(0.03)			(0.03)	(0.03)
Treated=1 # Guardian: 2-yr degree					-0.015			-0.008	-0.013
					(0.04)			(0.04)	(0.04)
Treated=1 # Guardian: 4-yr degree					-0.004			-0.013	-0.015
					(0.03)			(0.04)	(0.04)
Treated=1 # Guardian: Master's or higher					-0.001			-0.015	-0.014
					(0.04)			(0.04)	(0.04)
constant	0.693***	0.672***	0.701***	0.668***	0.671***	0.647***	0.648***	0.641***	0.640***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
R-squared	0.117	0.123	0.143	0.146	0.132	0.152	0.177	0.160	0.185
N	791	791	791	774	791	774	774	774	774

Standard errors in parentheses.

Table 5 – OLS regressions with treatment-covariate interaction terms.

Interacting the treatment dummy with demographic variables allows us to decompose overall score differences across groups into differences in their baseline levels of ad literacy and differences in their responses to treatment. The results in Table 5 echo those in Table 4 while adding some nuance. For example, Table 5 continues to show suggestive evidence that older and higher-income students perform better than others, all else equal. However, negative coefficients on the associated interaction terms also suggest that these groups receive less of a boost from Admongo than their younger/lower-income counterparts do. Thus, it may be that younger and/or lower-income students stand to gain the most from experiencing Admongo, in that these subjects exhibit the lowest level of ad literacy at baseline and experience the largest absolute improvement from treatment. Figure 7, below, shows the pattern along the grade-level margin. Treatment appears to erase grade-level advantages on test performance.

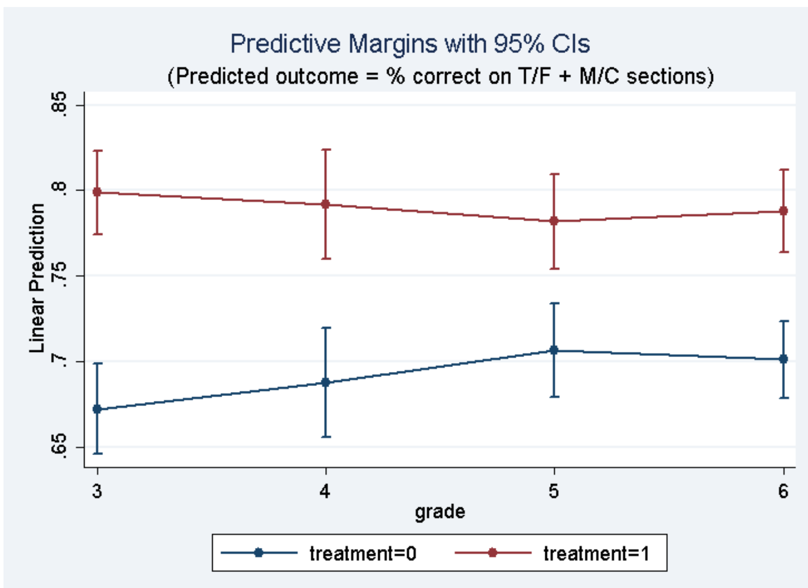


Figure 7. Admongo-grade-level interactions.

One possible explanation for the pattern in Figure 7 is that younger children are less experienced with marketing and therefore come to Admongo with more room for improvement. Another, though admittedly more speculative, is that the treatment itself – a relatively low-tech online game - is less engaging for older children who are more likely to be accustomed to immersive gaming experience of platforms like Sony PlayStation or Microsoft XBox. These children may pay less attention to the game and its messaging as a result.

Of course, as in Table 4, these results are only suggestive and not definitive. Only one of the across-grade-level, within-treatment-status contrasts in Figure 7 is statistically significant (control group, 5^{th} grade vs 3^{rd} grade, $p=0.0745$), meaning that one cannot say definitively what the age-related pattern is with or without treatment. Clearly, the bigger story here is the overall treatment effect, which in Table 5 ranges from 9 to 13 percentage points, depending on specification, and which is always significant at the 1% level.

C. Robustness

i. Propensity Score Matching Estimates of the Average Treatment Effect

In this section, as a robustness check, we estimate the effect of Admongo using a matching estimator. Matching could outperform regression in certain situations by excluding observations that lack counterparts in the opposite treatment group (Dehejia and Wahba, 2012). The lack-of-counterparts problem is a real concern in our context; as we have seen, the treatment regimen selected for a different mix of children than the control regimen (higher-income, fewer minorities, etc.).

To illustrate the concern, consider a regression-based estimate of the Admongo treatment effect on a hypothetical sample in which nearly all treatment subjects attended private school, were in the sixth grade, and came from upper-income households with highly educated parents, while nearly all control students attended public school, were in the third grade, and came from lower-income households headed by parents with a GED or less. Regression will attempt to "control" for these differences in baseline characteristics, but the two populations are also so different as to raise questions about the suitability of the comparison. The pervasive imbalances suggest that we are missing the counterfactual outcomes we need in order to disentangle treatment effect from the effects of other confounding differences across the two samples.

A matching-based approach would begin by narrowing the sample to control-treatment pairs with similar covariate values. The downside to this approach is that it might cut down the sample substantially and thereby narrow the interpretation of the estimand to a treatment effect at particular values of the covariates. However, by insuring a more like-with-like comparison, matching has the potential benefit of preventing over-generalization of results beyond the sub-populations for which we have adequate data.

We implement matching using the propensity score, which is just the probability of being treated, conditional on other covariates.[11] The potential benefits of using propensity score matching to increase balance have been noted by numerous researchers in the statistics and econometrics literatures. Austin (2011) cites several potential benefits, including: i) model fit is easier to evaluate in a matching model than a regression model; ii) the matching approach "separates design of the study from the analysis of the study", in the sense of providing a formal structure for building a comparison group, which is independent of outcome measures; and iii) matching forces the researcher to confront the degree of overlap in covariates between control and treatment in a way that regression does not.

Glynn et al. (2006) echo Austin in noting propensity score matching's formal framework for getting rid of outliers in the data. Glynn et al. also tout the fact that a matching approach based

[11] An alternative is to match directly on covariates values. While matching directly on the covariate is sometimes possible, doing so becomes increasingly infeasible as the number of covariates grows large. Furthermore, Rosenbaum and Rubin (1983) show that whenever matching directly on covariates solves the selection problem (i.e., controls for all outcome-relevant differences in the two populations), matching on the propensity score does so as well.

on the propensity score does not assume the outcome is linear in the covariates, as regression commonly does. Finally, Dehejia and Wahba (1999) and Deheija (2005) demonstrate that propensity score stratification and propensity score matching can improve on regression alone when control and treatment groups differ substantially in the covariates.

Each subject's propensity score is her predicted likelihood of being in the treatment group, and this is estimated via a logistic regression of the treatment dummy on covariates. Table A3, in the Appendix, presents the results from this regression. Using the coefficient vector β estimated from this regression, the propensity score for each subject i, given her covariate vector X_i, is: $\Pr(\text{treated}|X_i) = \beta X_i \in [0,1]$.

In Figure 8, below, we plot the distribution of estimated propensity scores for control and treatment groups in the raw and the matched data. By "matched" data, we mean the sample of control-treatment pairs constructed by nearest-neighbor matching on the propensity score.

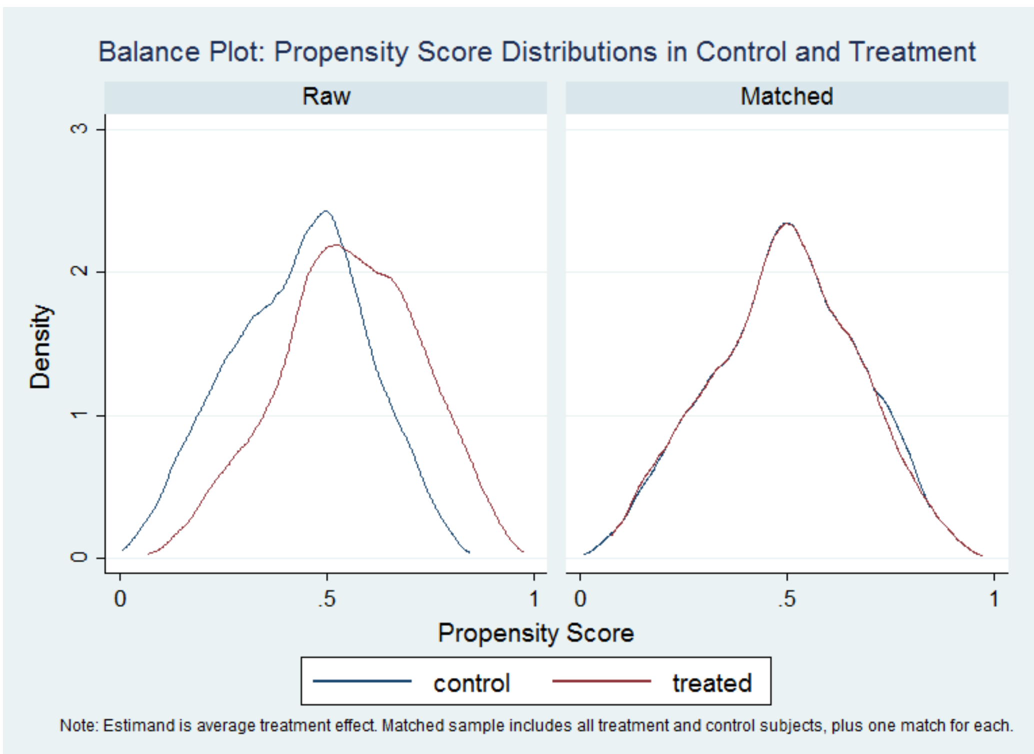


Figure 8. Raw-sample and matched-sample distributions of the propensity score.

The left-hand side of Figure 8 depicts the distribution of the propensity score by treatment status in the raw data, i.e. the distribution of estimated propensity scores for the original 774 control and treatment subjects with the required covariates. As one would expect given the differences observed earlier in Table 2, the propensity score distribution for control subjects is distinct from that of treatment subjects and sits, by construction, to the left of the treatment distribution. This outcome is simply a reflection of the fact that treatment and control differ in systematic and observable ways that the logit model detects.

The left-hand plot of Figure 8 also shows there is substantial overlap in the two distributions: the vast majority of the control group distribution falls within the range of propensity scores observed in the treatment group. Furthermore, both distributions are single-peaked near the center of the observed range of values. These observations imply a fairly high degree of comparability between control and treatment. Both groups are constituted in large part by subjects whose characteristics make them nearly equally likely to be found in either group.

The right-hand side of Figure 8 presents the analogous propensity score distributions, but for the matched data. Here the matched data consists of all n=774 original subjects, plus one propensity-score match for each, yielding a total of 1,548 observations. (The average treatment effect, estimated from this matched data, will be an average of the 774 individual contrasts.)

Control and treatment distributions in the matched data overlap each other more or less exactly, and their shape is a hybrid of the two raw data distributions from the left-hand side. The reasons underlying the differences between the left and right-hand side plots are i) we match with replacement, meaning the relative frequency of particular control-group observations can be different in the matched data, and ii) Figure 8 reflects estimation of the average treatment effect (ATE), which matches not only every treated observation, but also every control observation (thus altering the relative frequency of treatment observations as well).[12]

To determine whether our propensity score model is well-specified, we need to confirm that matching achieved balance in the covariates that are potentially relevant for test performance (Austin; 2009, 2011). Table 7, below, assesses control-treatment balance by comparing means and variances of covariates across the groups. The “standardized difference” equals treatment-group mean minus control-group mean, divided by average standard deviation (Austin, 2011). The “variance ratio” equals treatment-group variance divided by control-group variance. A standardized difference of zero and variance ratio of one represents the ideal of perfect balance, at least in terms of the first two moments of the data. We can see from Table 7 that matching on the propensity score has improved balance in the covariates, especially in terms of the means.

[12] The appendix shows the analogous density plots following estimation of the average treatment effect for the treated (ATET). Because estimation of ATET matches only treatment group observations, the matched-data density for both groups converges to the treatment group’s raw-data density.

Balancing of Covariates in the Raw and Matched Data

	Standardized differences		*Variance ratios*	
	Raw	**Matched**	**Raw**	**Matched**
$40K-$50K	-0.122	-0.056	0.732	0.859
$50K-$75K	0.115	-0.042	1.160	0.951
$75K-$100K	0.140	0.051	1.293	1.107
$100K+	0.166	0.047	1.204	1.054
only child	-0.152	-0.010	0.779	0.985
college grad	0.110	0.023	0.993	0.998
married	0.202	-0.021	0.793	1.022
full-time emp.	-0.110	0.042	1.041	0.990
South	-0.109	-0.039	0.929	0.971
white	0.203	-0.015	0.773	1.018
morning	-0.079	-0.044	0.925	0.955
afternoon	0.314	-0.003	1.347	0.998
evening	-0.042	0.035	0.975	1.024
homeschooled	0.146	-0.023	1.619	0.935
4th grade	-0.013	0.039	0.977	1.076
5th grade	-0.001	-0.006	0.998	0.992
6th grade	-0.083	-0.063	0.939	0.959
weekend	0.236	0.031	1.774	1.088
male	0.026	0.031	1.000	1.002
guardian 40+ yrs.	-0.050	0.023	1.005	0.996

Note: Raw data consists of n=774 (381 treated, 393 control). Matched data consists of n=1,548 (774 treated, 774 control). Estimand is ATE.

Table 7. Comparing covariate means and variances across groups, before and after propensity-score matching.

Further, the density plots in Figure 9, below, show that in general, matching on the propensity score does a good job of balancing control and treatment throughout a covariate's distribution, and not just on mean and variance.

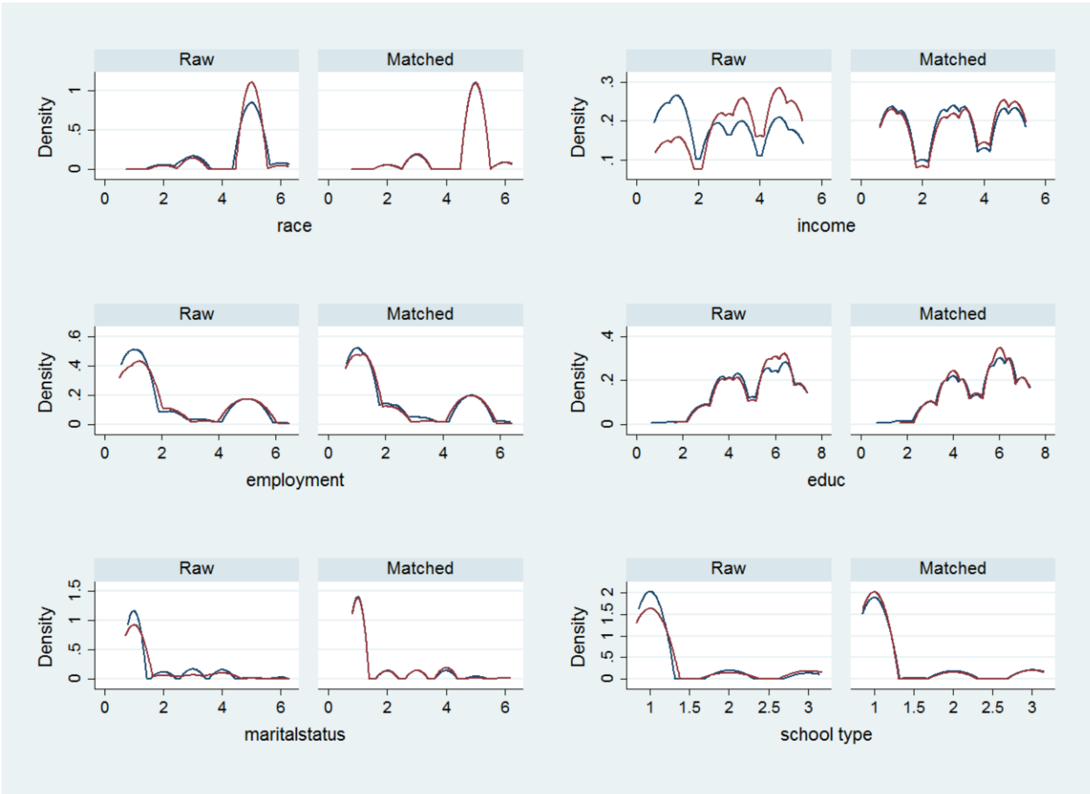


Figure 9. Covariate densities in raw vs. matched data (red = treated, blue=control).

Having demonstrated that propensity score matching does improve the balance of control and treatment in our data, we now present our matching-based treatment effect estimates. Table 8, below, presents estimates of both the average treatment effect (ATE) and average treatment effect on the treated (ATET) after matching on propensity scores. The former is a prediction for the entire sample, whereas the latter applies to the treated participants only. For ease of comparison to regression results, Table 8 also presents regression-based treatment-effect estimates that control for exactly the same covariate vector that was used in the propensity score model (see Table AX in Appendix).

Admongo Treatment Effect Estimates - Propensity Score Matching vs. Regression

A. Propensity score matching estimates.

	Coefficient	Abadie-Imbens robust std. err.	z-stat	P>|z|	95% c.i.		N
ATE	0.089	0.012	7.250	0.000	0.065	0.113	774
ATET	0.090	0.014	6.320	0.000	0.062	0.118	774

B. Regression estimates, controlling for same covariates as propensity score logit model.

	Coefficient	Robust std. err.	z-stat	P>|z|	95% c.i.		N
ATE	0.090	0.010	8.810	0.000	0.070	0.110	774
ATET	0.092	0.011	8.520	0.000	0.071	0.113	774

Table 8. Propensity-score matching estimates of Admongo treatment effect vs. regression-based estimates.

The conclusion that we take from Table 8 is that propensity-score based estimates of Admongo's effect are essentially identical to regression-based estimates. This statement is true whether we consider the ATE or ATET. Recall that regression-based estimates in Tables 4 and 5 were all in the 0.08-0.13 range, meaning the equivalence of regression and propensity score estimates is not contingent on *which* regression specification is used. Thus, we conclude that lack of covariate balance in control and treatment per se is not causing bias in the OLS estimates of Admongo's impact.

Matching and regression can potentially lead to different estimates because of the way the two methods weight observations differently.[13] However, the impact of Admongo across the range of covariate values appears to be uniform enough that the difference in weighting schemes between matching and regression has virtually no effect on estimates of the treatment effect. This result was foreshadowed by the test-score histograms in Figures 5 and 6, which demonstrated similar rightward shifts in the treated group's score distributions, across a broad range of demographic subgroups.

In summary, the use of propensity score matching to mitigate the impact of imbalances in control and treatment does not change the main take-away from the OLS regressions: playing the Admongo game leads to, on average, approximately a nine percentage-point increase in score on the ad literacy exam.

ii. Intent-to-treat analysis: a conservative approach to measuring Admongo's impact

The treatment-effect estimates presented above may be interpreted as causal effects of Admongo only under the assumption that, after controlling for covariates, treatment status is independent of a subject's potential outcomes (Rosenbaum and Rubin, 1983; Rubin, 1974). When this assumption is satisfied, control and treatment subjects at a given covariate value are draws from the same distribution and can serve as valid counterfactual outcomes for each other. We have maintained this "conditional independence" assumption throughout the presentation of regression-based and propensity-score-based estimates. However, we cannot prove that it is satisfied.

There are many child-level and household-level characteristics that we do not observe in our data. Some of the unobserved characteristics could be both outcome-relevant and correlated with treatment status, confounding our ability to measure Admongo's effect in an unbiased way if we rely exclusively on the completer sample.

For instance, the time and effort required to play the Admongo game to the requisite end point may have selected for subjects with higher potential outcomes under treatment (compared to control-group counterparts, who were not asked to play the game). It could be that children willing to play the game on a voluntary basis were also more likely to absorb the instructional

[13] Angrist and Pischke (2009) show that matching weights the treatment effects observed at different covariate values in proportion to the probability of treatment at those values. In contrast, regression weights those covariate-specific effects in proportion to the variance in the treatment variable at those values. Thus, matching estimates lean most heavily on treatment effects observed at values of X where the probability of treatment is closest to one, whereas regression leans most heavily on treatment effects observed at values of X where the probability of treatment is closest to one-half.

material than children unwilling to play. Such "patient" or "receptive" children may tend to experience a higher boost from treatment. This higher potential benefit may not be captured by any of the observable controls and therefore may constitute an unmeasured confounding variable. Our treatment effect estimates in this case would overstate the impact of the Admongo game because they would inadvertently capture the combined effect of Admongo and unmeasured potential.

An alternative path that takes a more conservative approach is an intent-to-treat (ITT) estimator. In ITT, the objective is to estimate Admongo's impact using all randomized subjects - completers *and* quitters. The logic of ITT is to characterize an intervention's impact in a way that factors in real-world complications faced during implementation, such as subjects switching protocols (e.g. control assignees somehow acquiring the treatment), or treatment assignees avoiding some or all of the treatment (Gupta, 2011).

The ITT is not directly comparable to the ATE because it addresses a slightly different question. In the context of Admongo, the ATE approach asks, "what is the effect of playing Admongo?" whereas the ITT approach asks, "what is the effect of assigning someone to play Admongo?" If instructing a child to play Admongo always resulted in her playing the game (and if no child played Admongo who *wasn't* told to), then the ITT and ATE approaches would produce the same estimate.

An ITT analysis in our case would include all 4,203 recruits – the 791 completers and 3,439 quitters. However, because tests scores do not exist for quitters in our data, we need a way to predict how the quitters would have scored, had they taken the test. The medical literature offers different solutions to the problem of missing outcome data in ITT analyses, with suggestions ranging from multiple imputation, to complete-case analysis, to carrying forward the last observed outcome before drop-out (Alshurafa, 2012).

We opt for a simple bounding exercise in which we assume that treatment quitters would have performed, on average, the same as control completers. Put differently, we assume that treatment quitters completed the required gameplay but would have derived no benefit in terms of test performance. If one is willing to rule out a negative treatment effect, then this approach can be thought of as maximally conservative. Under this assumption, a back-of-the-envelope ITT effect estimate is then just the weighted average of Admongo's impact on the treatment quitters (zero, by assumption) and its impact on the treatment completers ($\approx$ 0.09, according to regression and propensity score estimates): $\frac{390}{3591} \times 0.09 + \frac{3201}{3591} \times 0 = 0.0098$, or about one percentage point. Note, again, that if any significant fraction of treatment quitters would have derived a positive benefit from Admongo, then this one percentage-point estimate will have underestimated Admongo's true effect. It is useful as an extreme lower bound.

The considerably lower one percentage-point figure primarily reflects the low rate of participation among children assigned to play Admongo. What should we make of such a low level of compliance? What does it say about Admongo's likely impact on its target audience, and are there broader lessons for educational outreach programs in general.

While we cannot extrapolate in any formal way from our non-random sample of 8-12 year olds, the treatment group's 89% attrition rate does suggest room for improvement in the game's appeal to its target audience. Having said that, we also should keep in mind the context under which a treatment assignee decided whether to quit in our experiment before we draw inferences about the game's appeal to children in a non-experimental setting. Panel members periodically have the opportunity to participate in marketing studies in return for compensation, meaning time spent playing Admongo is, to some extent, being weighed against other potentially less time-consuming and/or less taxing opportunities for earning compensation. It is possible that this framing results in lower participation than would be observed in a classroom setting, where a teacher suggests that students use their free time to play Admongo – and lower, still, compared to a classroom setting where Admongo was assigned for a grade.

V. Discussion and Conclusion

In this paper, we have analyzed results from a randomized experiment that evaluated the impact of an FTC campaign targeted at children in the third through the sixth grade. Our analysis shows that playing the Admongo game for a relatively short time is associated with a score increase of approximately nine percentage-points on an ad literacy test. This effect is the equivalent of roughly two additional correct answers on a 25-question test. The magnitude of the estimate is relatively stable across demographic groups, robust to the inclusion of demographic controls in OLS regressions, and robust to the use of an alternative treatment-effect estimator (propensity score matching).

Under the assumption of conditional independence – i.e., that potential outcomes of control and treatment participants are equivalent, once we control for observed covariates – the above estimates can be regarded as causal treatment effects. However, the disproportionately high rate of attrition in the treatment arm (81%, vs. 37% in control) counsels caution in this regard. Participants in the treatment group evidently were filtered in a way that control participants were not. We cannot say that this selection mechanism definitely resulted in unmeasured and outcome-relevant differences between groups, but neither can we rule it out. For this reason, it is possible that our ATE estimates overstate the impact of Admongo on test performance. Our quasi-ITT approach provided a more-conservative estimate of a 1% effect. This conservative approach assumed that all treatment-quitters would have derived *zero* benefit from the game. This is a stark, and in our view unrealistic, assumption - but useful, at least, for putting a floor under possible effect sizes.

The experiment's reliance on a marketing research panel, the voluntary nature of participation, the high attrition rate in the treatment group, and finally, the lifting of stratification quotas during fielding constrain our ability to generalize to the wider population of U.S. 8-12 year olds. The particular quiz employed to measure ad literacy also likely played a role in the measured effect size. In short, the true expected effect of Admongo on the advertising literacy of a randomly selected population of U.S. 8-12 year olds is not discernible from our experiment. Still, our results do suggest the potential of a relatively quick and playful intervention to enhance children's advertising literacy in ways that conceivably improve their critical thinking and decision making in commercial contexts.

Our experiment did not measure long-term outcomes; and while they are conceivable, we suspect that a more lengthy intervention would be required in order to cause prolonged changes in the way children process advertising and make purchase decisions. Additionally, a considerably more involved and more expensive evaluation would be necessary to measure long-term effects, requiring follow-up with subjects over a prolonged period (say, multiple check-ins over a series of months) and the recording of reactions, thoughts and purchase behavior over time in journals. It would be interesting to compare the preliminary evidence of a treatment effect found in this paper to the results of a longer-term randomized study of a more comprehensive intervention, such as the full Admongo curriculum, carried out over a series of weeks in 3rd-6th grade classrooms.

VI. Appendix

Table A1. Table A1, below, presents frequencies of select characteristics among the four sub-populations in the data: control quitters (Assigned Control and quit), control completers (Assigned Control and completed), treatment quitters (Assigned Treatment and quit) and treatment completers (Assigned Treatment and completed).

Select Covariates, by Assignment and Completion Status

	Assigned Control		*Assigned Treatment*	
	Quit*	Completed	Quit	Completed
N	238	401	3,201	390
Over $50K	65%	55%	73%	72%
White	63%	72%	76%	80%
Married	74%	70%	77%	78%
2+ siblings	32%	31%	33%	39%
Over $100K	26%	22%	34%	29%
Bachelor's +	55%	49%	59%	55%
Single	8%	10%	6%	6%
Black	9%	15%	9%	11%
South	38%	37%	33%	32%
FT employed	57%	62%	69%	56%
Only child	31%	22%	21%	15%
Under $40K	20%	30%	14%	17%
homemaker	23%	21%	16%	24%

**Covariate percentages for the control quitters were estimated using the N=65 sub-sample for which these variables are observed.*

Table A1.

Table A2. Table A2, below, shows that lifting the income quotas after June 1 did lead to a modest increase in the concentration of higher-income subjects entering the sample. However, the net effect of lifting quotas on covariate balance between groups was modest for three reasons.

First, prior to lifting quotas, the treatment group already lagged behind the control group in attracting lower-income subjects, with just 23% of its under-$40K income quota met as of May 16 (compared to 79% met in the control group). Relaxing the income quotas simply dropped a target that was not being met in the field. Second, the income profile of the treatment group changed only slightly following the lifting of income quotas, with the percentage of households earning $100K+ rising from 29% to just 32%, and with the percentage earning $50K+ actually falling. Third, the influx of more higher-income students into the control group post-quotas had little impact on that group's overall composition because 385 of its 400 subjects (96%) had been recruited already by the time quotas were lifted.

Characteristics of Quitters and Completers, Before and After Sampling Income Quotas Were Lifted, June 1, 2016

	Household Characteristics					Child Characteristics				Parent Characteristics				
	N	income < $40K	income >$50K	income >$100K	Census region South	white	black	Only child	2+ siblings	Parent married	Parent single	Parent Bachelor's +	Parent full-time employed	Parent home-maker
	While income quotas were in place (March 28-June 1)													
control quits*	209	21%	63%	26%	13%	66%	9%	24%	19%	73%	8%	56%	57%	24%
control completes	385	31%	55%	22%	36%	72%	15%	22%	32%	70%	10%	49%	62%	21%
treatment quits	2,579	16%	73%	33%	34%	77%	9%	22%	32%	75%	7%	58%	68%	16%
treatment completes	342	17%	74%	29%	32%	82%	9%	16%	39%	79%	6%	54%	57%	22%
	After income quotas were lifted (June 2-24)													
control quits*	29	0%	100%	25%	10%	60%	0%	18%	14%	100%	0%	50%	67%	0%
control completes	16	13%	87%	33%	56%	69%	0%	13%	25%	69%	6%	56%	69%	6%
treatment quits	622	8%	87%	44%	28%	72%	10%	19%	38%	82%	6%	66%	70%	16%
treatment completes	48	17%	71%	32%	25%	71%	21%	10%	40%	77%	8%	58%	44%	35%

**Covariate percentages for the control quitters were estimated using the N=65 sub-sample for which these variables are observed.*

***Participants categorized using variable "startdate"*

Table A2. Comparison of participants who enrolled before and after income quotas were lifted.

Table A3. Table A3, below, presents results of a logistic regression of the treatment dummy on covariates. This regression is used to predict the probability that subject *i* is treated, conditional on covariate values X_i . This probability is subject *i*'s propensity score.

Propensity Score Matching
Treatment Model

logistic regression

dependent variable = Admongo treatment dummy

(z-stats in parentheses)	
$40K-$50K	0.431
	(1.48)
$50K-$75K	1.099***
	(4.35)
$75K-$100K	1.334***
	(4.6)
over $100K	1.263***
	(4.47)
only child	-0.149
	(-0.71)
college grad	-0.007
	(-0.04)
married	-0.133
	(-0.63)
full-time emp.	-0.433*
	(-2.51)
South	-0.133
	(-0.79)
white	0.388*
	(2.03)
morning	1.269***
	(3.59)
afternoon	1.873***
	(5.32)
evening	1.345***
	(3.9)
homeschooled	0.738*
	(2.33)
4th grade	-0.113
	(-0.47)
5th grade	-0.06
	(-0.27)
6th grade	-0.246
	(-1.20)
weekend test taker	0.980***
	(3.79)
male	-0.048
	(-0.31)
guardian 40+ yrs.	-0.331*
	(-1.99)
constant	-1.979***
	(-4.59)
N	774
pseudo R-squared	0.0952
chi-squared	102
p-value	0.000

** p<0.05, ** p<0.01, *** p<0.001*

Table A3. Propensity score matching, stage one: logistic regression of Admongo treatment dummy on covariates. Sample = completers sample.

Figure A1. Figure A1, below displays the propensity score density plots from an estimate of the average treatment on the treated (ATET), meaning the matched sample comprises the 381 original treated subjects plus one propensity-score match from the control group for each. Matching is with replacement. The raw data distributions in the left-hand plot are identical to those in Figure 9 where estimation was of ATE. Control and treatment distributions converge in the matched sample, as they did in Figure 9. However, unlike in Figure 9, the treated distribution is unchanged going from raw to matched, and the control distribution converges to the treated distribution. This pattern reflects the one-way matching of the ATET estimation: all treated are matched to a control observation, but not vice versa.

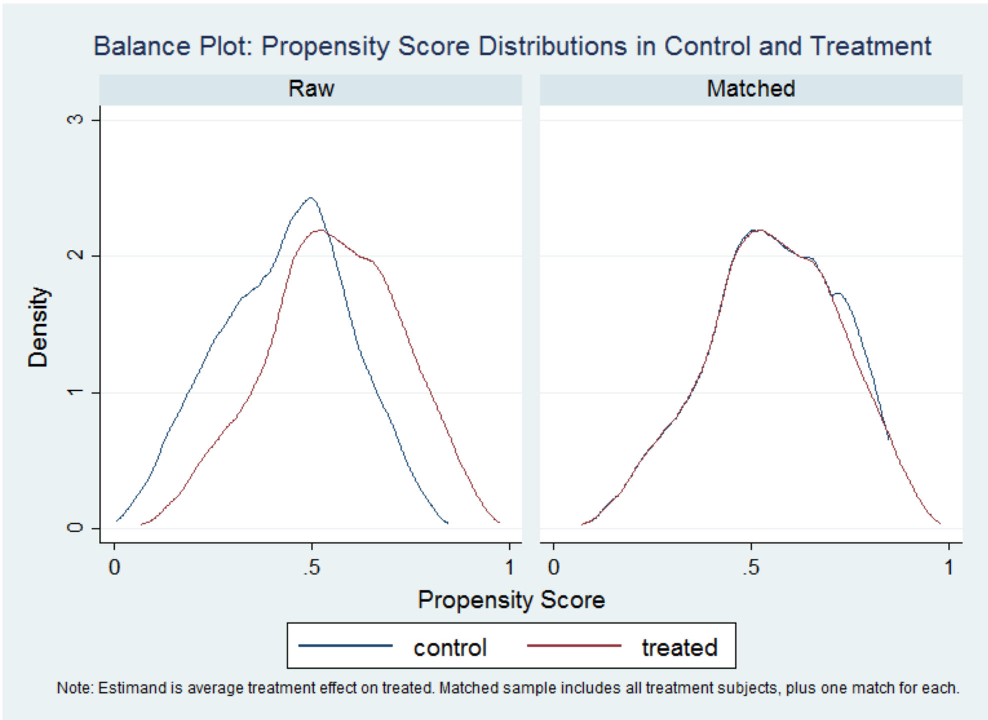


Figure A1. Propensity score distributions following estimation of ATET.

Figure A2. Graphical timeline of sample recruitment. Recruitment of the 400 control completers proceeded much more quickly than recruitment of the 400 treatment completers due to heavy attrition among treatment assignees.

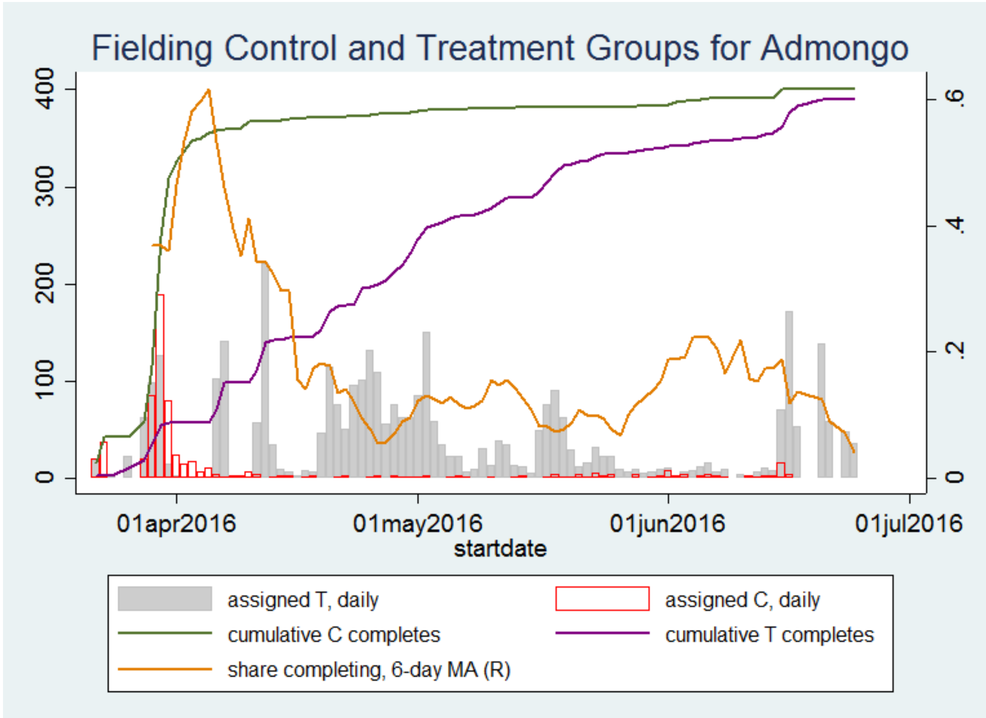


Figure A2. – Admongo evaluation recruitment process over time.

Figure A3. The ad literacy test taken by participants.

AD LITERACY TEST

PART 1: TRUE/FALSE

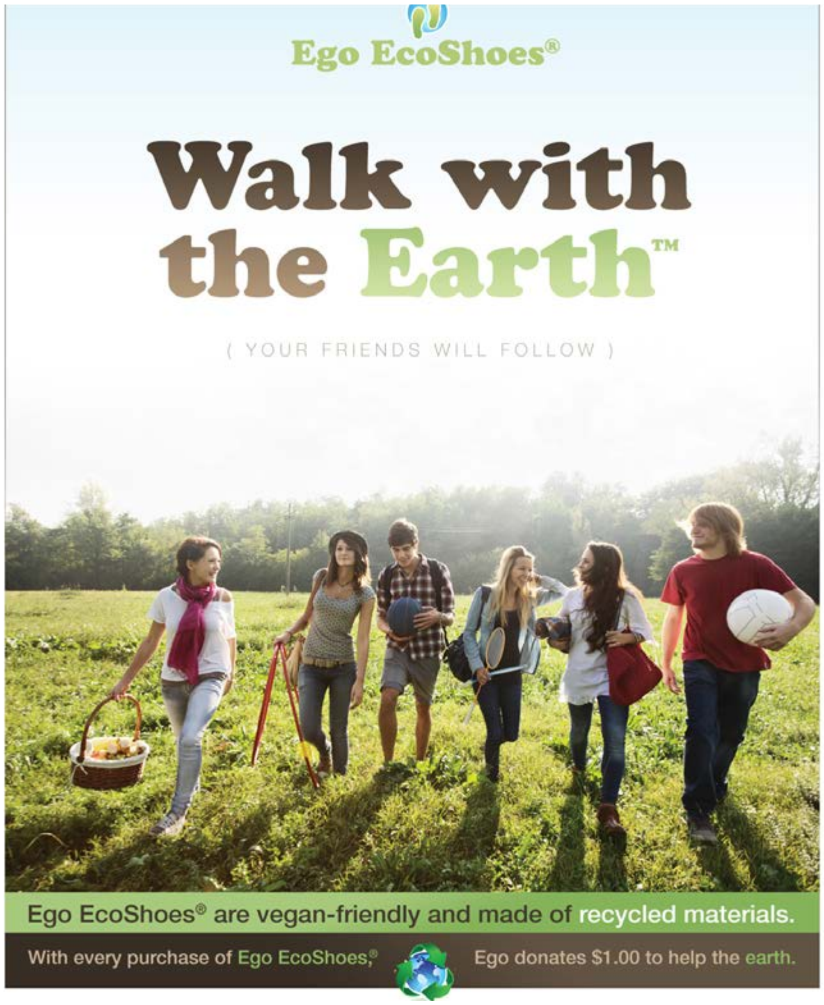
There are 13 statements in this section. After reading each statement, please click on whether you believe it is true or false. Please include only one answer per line.

	STATEMENT	*True*	*False*
1.	Advertising can only tell you about products. It cannot try to convince you to buy something.		
2.	Advertising helps pay for websites and magazines.		
3.	The government has to approve an ad before you can see it.		
4.	Advertisements have to tell the truth about how much you will like a product.		
5.	The “target audience” for an advertisement can be a website, a TV show, a billboard, or a radio program.		
6.	Advertisers can only direct advertisements to people over the age of 12.		
7.	When you go to a website for a toy or candy, the games you play at that		

	site are a kind of advertising.		
8.	Wearing a brand's logo on your t-shirt is a kind of advertising.		
9.	Advertisements are not allowed to look like magazine articles.		
10.	It is possible for advertisers track the sites you visit on the internet.		
11.	By law, all advertisements have to tell you any negative information about a product.		
12.	When a basketball player appears in an advertisement for sneakers, that player has been paid to wear that brand of shoes.		
13.	Funny ads use humor to try to give you a good feeling about the product being advertised		

Thanks for answering those questions. Click here to move on to the next section.

PART 2: MULTIPLE CHOICE
In this section you will be looking at three different ads. After you look at each of these ads, you will be asked to answer four multiple-choice questions.



Questions 14-17 are about Advertisement #1, above.

14. What is the target audience of this advertisement?

a. Two boys and four girls
b. Walk with the Earth™ (Your Friends Will Follow)
c. Teenagers and young people.
d. "Ego EcoShoes® are vegan-friendly and made of recycled materials."

15. How will buying Ego EcoShoes help the earth?

a. With every purchase of EcoShoes, Ego donates $1 to the *Walk With The Earth™* foundation.
b. With every purchase of EcoShoes, Ego donates $1 to vegan-friendly organizations.
c. With every purchase of EcoShoes, Ego donates $1 to support recycling programs.
d. It's not clear how buying Ego EcoShoes will help the earth.

16. What is one way the advertisement tries to convince you to buy EcoShoes?

a. It shows a group of famous people wearing EcoShoes.
b. It explains why EcoShoes are comfortable for long walks outside.
c. It connects EcoShoes with having fun and helping the environment.
d. It uses humor about vegetarians and recycling.

17. What does the EcoShoes advertisement really want you to do?

a. Spend time outside with your friends.
b. Buy a pair of EcoShoes.
c. Try to use more recycled materials.
d. Donate $1 to help the Earth.

Thanks so much for your work so far. Click here to move on to the next section.

Questions 18-21 are about Ad #2, above.

18. What do the <u>words</u> in the advertisement say about Holobox?

a. It costs more than some other gaming systems because it has the latest hologram technology.
b. It has 32 Gigabytes of SDDR7 memory and costs $399.
c. Its 16 core processor allows faster play speeds than other gaming systems.
d. It has dual band Wireless Z capabilities that allow you to play multi-player games.

19. What does the advertisement say is included in the $399 price of the product?

a. Two free games, and the gaming console.
b. One controller, two free games, and the gaming console.
c. One controller, two free games, a wireless z connection, and the gaming console.
d. It doesn't say what is included.

20. Where would you most expect to see this advertisement?

a. On a website that features cheat codes for popular video games.
b. On a commercial during a show about a high school glee club.
c. On a billboard near a large children's hospital.
d. On the scoreboard at a baseball game.

21. Why does the boy in the advertisement look so interested in playing the game?

a. It's the first time that he has played Holobox, and he can't believe how fast and realistic it is.
b. The advertiser wants you to believe you'll have a good time playing Holobox.
c. Free Holobox games like HOLO-PET and HOLO-BALL are interesting to everyone.
d. Most teenagers are curious about new technologies.

Questions 22-25 are about Ad #3, above.

22. What is the target audience of this advertisement?

a. Young women who want to look better.
b. The cast of the hit TV show *Surfer High*
c. "You won't believe the difference!"
d. MetaLean Yogurt, Arctic Berries flavor.

23. What do the words in the advertisement say?

a. In order to get the results listed, you need regular exercise.
b. In order to get long, lean muscles you just need to drink this product.
c. Drinking arctic berry yogurt is the healthiest way to get calcium and protein.
d. Arctic berries give you energy.

24. Why does the advertisement mention the show *Surfer High*?

a. The girl in the advertisement must have gone to the school on *Surfer High*.
b. The advertiser thinks you will buy a drink that *Surfer High* cast members like.
c. *Surfer High* has a workout program that builds long, lean muscles.
d. Research shows that high school students like yogurt-based smoothies.

25. Which of these does the MetaLean advertisement *not* use to get you to buy it?

a. Makes a claim about MetaLean's health benefits.
b. Offers prizes, sweepstakes, and gifts.
c. Features a special ingredient.
d. Includes a testimonial or endorsement.

Thank you for looking at that. Please click here to move on to the last section.

PART 3: FIND THE ADS

In this section, you will be looking at a drawing. Click on everything in the drawing that you believe is an advertisement.

26. For this question, click on everything that you believe is an ad in this drawing. After doing that, count up all the items you've clicked on, and answer the question, How many ads are in this picture?

a. There are **XX** ads in this picture
b. There are **XX** ads in this picture
c. There are **XX** ads in this picture.
d. There are **XX** ads in this picture.

PART 4: OPINION QUESTIONS

Questions 27-35 are opinion questions. That means, they are about what *you* think and do. <u>There are no right or wrong answers to questions 27-35</u>. Just select the answer that best matches what you believe, or how you behave.

27. How often do you ask your friends' whether they like a product before you decide if you want to get it?

a. All the time
b. Some of the time

c. Hardly ever
d. Never

28. The advertisements that get my attention the most are the ones that:

a. Show a famous person that I like
b. Show a cartoon character that I like
c. Are funny
d. Have a cool song

29. When I watch TV, I mostly watch

a. Cartoons
b. Shows with real kids my age or a little older
c. The shows my parents or older brothers and sisters watch
d. A little bit of everything

30. I learn about new things I want mostly from

a. TV
b. The internet
c. My friends and family
d. Lots of different places

	STATEMENT	**Yes**	**No**
31.	Advertisements usually tell me the most important information about a product.		
32.	When advertisements say something is the “best” or “greatest,” I tend to believe it.		
33.	I’ve clicked on an online ad because it promised me I’d get something free.		
34.	My parents have talked to me about being tricked by what I see in advertisements.		
35.	In the past, I have bought a product that disappointed me because it didn’t do what it said it would in the advertisement.		

Thank you very much for participating in this project.

BIBLIOGRAPHY.

A. Smith, Jeffrey & E. Todd, Petra (2005). "Does matching overcome LaLonde's critique of nonexperimental estimators?" *Journal of Econometrics*, Elsevier, vol. 125(1-2), pages 305-353.

Alshurafa M, Briel M, Akl EA, Haines T, Moayyedi P, Gentles SJ, et al. (2012). Inconsistent Definitions for Intention-To-Treat in Relation to Missing Outcome Data: Systematic Review of the Methods Literature. *PLoS ONE* 7(11): e49163.

An, Soontae, Hyun Seung Jin and Eun Hae Park (2014). "Children's Advertising Literacy for Advergames: Perception of the Game as Advertising." *Journal of Advertising*. v. 43, Issue 1, pp 63-72.

Andronikidis, A. I. and Lambrianidou, M. (2010). Children's understanding of television advertising: A grounded theory approach" *Psychology & Marketing*, 27: 299-322.

Angrist, Joshua D. and Jorn-Steffen Pischke, (2009). *Mostly Harmless Econometrics.* Princeton University Press, Princeton, New Jersey.

Austin, Peter C. 2009, "Balance Diagnostics for Comparing the Distribution of Baseline Covariates between Treatment Groups in Propensity-Score Matched Samples." *Statistics in Medicine* 28.25: 3083–3107.

Austin, Peter C. (2011). "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate Behavioral Research* 46.3: 399–424.

Bagwell, Kyle and Gea M. Lee, (2010). "Advertising Competition in Retail Markets" *The B.E. Journal of Economic Analysis& Policy.* Vol. 10, Issue 1, Article 70.

Beales, J. Howard (2004). "Advertising to Kids and the FTC: A Regulatory Retrospective That Advises the Present". Available at: <https://www.ftc.gov/sites/default/files/documents/public_statements/advertising-kids-and-ftc-regulatory-retrospective-advises-present/040802adstokids.pdf>

Benham, Lee. (1972). "The Effect of Advertising on the Price of Eyeglasses." *The Journal of Law & Economics*, 15(2), pp. 337-352.

Calfee, John E., & Ringold, D. (1994). The 70% Majority: Enduring Consumer Beliefs about Advertising. *Journal of Public Policy & Marketing, 13*(2), 228-238.

Calfee, John E., & Janis K. Pappalardo, (1991). "Public Policy Issues in Health Claims for Foods." *Journal of Public Policy & Marketing.* Vol. 10, No. 1, pp. 33-53.

Chester, Jeff (2012). "Kids Spending and Influencing Power: $1.2 Trillion says leading ad firm". Center for Digital Democracy. November 1, 2012

Donohue et al., (1980). "Do Kids Know What TV Commercials Intend?" *Journal of Advertising Research. Vol. 20, no. 5.*

Esther Rozendaal, Matthew A. Lapierre, Eva A. van Reijmersdal & Moniek Buijzen, (2011). "Reconsidering Advertising Literacy as a Defense Against Advertising Effects", *Media Psychology*, 14:4, 333-354

Federal Trade Commission (2007). *Children's Exposure to TV Advertising in 1977 and 2004*: *Information for the Obesity Debate.* Available here: https://www.ftc.gov/sites/default/files/documents/reports/childrens-exposure-television-advertising-1977-and-2004-information-obesity-debate-bureau-economics/cabecolor.pdf

Federal Trade Commission (1999). *Self-Regulation in the Alcohol Industry: A Review of Industry Efforts to Avoid Promoting Alcohol to Underage Consumers.* Available here: https://www.ftc.gov/sites/default/files/documents/reports/self-regulation-alcohol-industry-federal-trade-commission-report-congress/1999_alcohol_report.pdf

Federal Trade Commission (2003). *Alcohol Marketing and Advertising: A Report to Congress.* Available here: https://www.ftc.gov/sites/default/files/documents/reports/alcohol-marketing-and-advertising-federal-trade-commission-report-congress-september-2003/alcohol08report.pdf

Federal Trade Commission (2008). *Self-Regulation in the Alcohol Industry: Report of the Federal Trade Commission.* Available here: https://www.ftc.gov/sites/default/files/documents/reports/self-regulation-alcohol-industry-report-federal-trade-commission/080626alcoholreport.pdf

Federal Trade Commission (2014). *Self-Regulation in the Alcohol Industry: Report of the Federal Trade Commission.* Available here: https://www.ftc.gov/system/files/documents/reports/self-regulation-alcohol-industry-report-federal-trade-commission/140320alcoholreport.pdf

Federal Trade Commission (2008). *Marketing Food to Children and Adolescents: A Review of Industry Expenditures, Activities, and Self-Regulation.* Available here: https://www.ftc.gov/sites/default/files/documents/reports/marketing-food-children-and-adolescents-review-industry-expenditures-activities-and-self-regulation/p064504foodmktingreport.pdf

Federal Trade Commission (2012). *A Review of Food Marketing to Children and Adolescents: Follow-Up Report.* Available here: https://www.ftc.gov/sites/default/files/documents/reports/review-food-marketing-children-and-adolescents-follow-report/121221foodmarketingreport.pdf

Federal Trade Commission (2000). *Marketing Violent Entertainment to Children: A Review of*

Self-Regulation and Industry Practices in the Motion Picture, Music Recording & Electronic Game Industries. Available here: https://www.ftc.gov/sites/default/files/documents/reports/marketing-violent-entertainment-children/vioreport_0.pdf

Federal Trade Commission (2009). *Marketing Violent Entertainment to Children: A Sixth Follow-up Review of Industry Practices in the Motion Picture, Music Recording & Electronic Game Industries.* Available here: https://www.ftc.gov/sites/default/files/documents/reports/marketing-violent-entertainment-children-sixth-follow-review-industry-practices-motion-picture-music/p994511violententertainment.pdf

Fromm, Jeff (2018). "How Much Financial Influence Does Gen Z Have?" *Forbes*. Jan. 10, 2018.

Glynn, R. J., Schneeweiss, S. and Stürmer, T., (2006). "Indications for Propensity Scores and Review of their Use in Pharmacoepidemiology", *Basic & Clinical Pharmacology & Toxicology*, 98: 253-259.

Gupta, Sandeep K. "Intention-to-Treat Concept: A Review." *Perspectives in Clinical Research* 2.3 (2011): 109–112. *PMC*. Web. 19 May 2018.

Harris Interactive (2006). "Today's Youth Look to Advertising as Much as Their Friends When Making Purchase Decisions"

Hobbs, R. (2004). "Does Media Literacy Work? An Empirical Study of Learning How to Analyze Advertisements." *Advertising and Society Review*. V. 5. Issue 4;

Ippolito, Pauline M. and Alan D. Mathios. (1989). *Health Claims in Advertising and Labeling: A Study of the Cereal Market.* Bureau of Economics Staff Report, Federal Trade Commission, Washington, DC.

Ippolito, Pauline M. and Alan D. Mathios. (1990). "Information, Advertising and Health Choices: A Study of the Cereal Market." *The RAND Journal of Economics.* Vol. 21, No. 3, pp. 459-480

Ippolito, Pauline M. and Alan D. Mathios. (1995). "Information and Advertising: The Case of Fat Consumption in the United States" *The American Economic Review*, Vol. 85, No. 2, Papers and Proceedings of the Hundredth and Seventh Annual Meeting of the American Economic Association, pp. 91-95

James J. Heckman, Hidehiko Ichimura, Petra Todd, (1998). "Matching As An Econometric Evaluation Estimator", *The Review of Economic Studies*, Volume 65, Issue 2, Pages 261–294

Kunkel, D., B. L. Wilcox, J. Cantor, E. Palmer, S. Linn, and P. Dorwick. (2004). "REPORT OF THE APA TASK FORCE ON ADVERTISING AND CHILDREN, Section: Psychological Issues in the Increasing Commercialization of Childhood"

Lapierre, Matthew A., Frances Fleming Milici, Esther Rozendaal, Anna R. McAlister and Jessica Castonguay (2017). "The Effect of Advertising on Children and Adolescents." *Pediatrics.* V. 140, No. s2. November 2017.

Levin, Stephen R., et al. (1982). "Preschoolers' Awareness of Television Advertising." *Child Development*, vol. 53, no. 4, pp. 933–937.

Livingstone, S. and E. J. Helsper. (2006). "Does Advertising Literacy Mediate the Effects of Advertising on Children? A Critical Examination of Two Linked Research Literatures in Relation to Obesity and Food Choice." *Journal of Communication*. V. 56. 560-584

Martin, Mary C. (1997) "Children's Understanding of the Intent of Advertising: A Meta-Analysis." *Journal of Public Policy & Marketing*, vol. 16, no. 2, pp. 205–216.

Mesa, A. (2005). *Marketing and Tweens: BFF*, BrandChannel.com (Oct. 10, 2005).

Milyo, Jeffre., & Joel Waldfogel. (1999) "The Effect of Price Advertising on Prices: Evidence in the Wake of 44 Liquormart." *The American Economic Review*, *89*(5), pp. 1081-1096.

National Retail Federation (2019). "Keeping Up with Gen Z: Fall 2019 Consumer View." Available here: https://nrf.com/research/consumer-view-fall-2019

Nelson, Philip. (1974) "The economic value of advertising." In Yale Brozen (ed.), *Advertising and society*, pp. 43-66. New York: New York University Press.

Oates, C., M. Blades, B. Gunter and J. Don. (2003) "Children's understanding of television advertising: a qualitative approach." *Journal of Marketing Communications*. V. 9. 59-71.

Opree S., Rozendaal E. (2015). "The Advertising Literacy of Primary School Aged Children", In: Banks I., De Pelsmacker P., Okazaki S. (eds) *Advances in Advertising Research (Vol. V).* European Advertising Academy, vol 5. Springer Gabler, Wiesbaden

Pine, K., & Veasey, T. (2003). "Conceptualising and assessing young children's knowledge of television advertising within a framework of implicit and explicit knowledge", *Journal of Marketing Management*, *19*(3-4), 459-473.

Rajeev Dehejia, (2005). "Practical propensity score matching: a reply to Smith and Todd" *Journal of Econometrics*, Volume 125, Issues 1–2, Pages 355-364

Rajeev H. Dehejia & Sadek Wahba. (2012). "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, 94:448, 1053-1062.

Reijmersdal, Eva & Rozendaal, Esther & Buijzen, Moniek. (2012) "Effects of Prominence, Involvement, and Persuasion Knowledge on Children's Cognitive and Affective Responses to Advergames", *Journal of Interactive Marketing*, 26. 33-42.

Rosenbaum, Paul R. and Donald B. Rubin. (1983). “The Central Role of the Propensity Score in Observational Studies for Causal Effects” *Biometrika.* Vol. 70, No. 1. pp. 41-55.

Rozendaal E, Buijs L, van Reijmersdal EA., (2016) “Strengthening Children’s Advertising Defenses: The Effects of Forewarning of Commercial and Manipulative Intent”, *Frontiers in Psychology*.

Rubin, D. B. (1974). “Estimating causal effects of treatments in randomized and nonrandomized studies”, *Journal of Educational Psychology*, *66*(5), 688-701

StataCorp. (2017). *Stata 15 Base Reference Manual*. College Station, TX: Stata Press.

Stigler, George J. (1961) "The Economics of Information," Journal of Political Economy 69, no. 3: 213-225.