I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12158

When Paywall Goes AWOL: The Demand for Open Access Education Research

Seth Gershenson
Morgan S. Polikoff
Rui Wang

FEBRUARY 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12158

When Paywall Goes AWOL: The Demand for Open Access Education Research

Seth Gershenson
American University and IZA

Morgan S. Polikoff
University of Southern California

Rui Wang
American University

FEBRUARY 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

When Paywall Goes AWOL: The Demand for Open Access Education Research

As universities cut library funding and forego expensive journal subscriptions, many academic organizations and researchers, including the American Educational Research Association (AERA), are moving towards open-access publications that are freely downloadable by anyone with a working internet connection. However, the impact of paywalls on the consumption of academic articles is unclear. We provide novel evidence on this question by exploiting a natural experiment in which six high-impact, usually gated AERA journals became open access for a two-month period in 2017. Using monthly download data, and an always-open access journal as a control group, we show that making journals open access increased article downloads in those journals by 60 to 80% per month. Given a per-article download price of $36, this suggests a download elasticity of about 0.3 to 0.4.

JEL Classification: L17, O33
Keywords: open access, academic journal

Corresponding author:
Seth Gershenson
American University
School of Public Affairs
4400 Massachusetts Avenue NW
Washington, DC 20016
USA
E-mail: gershens@american.edu

Introduction

Academics, university libraries, and the general public are becoming increasingly frustrated by the rising costs of access to academic journal articles, particularly those published by for-profit firms (Bergstrom, 2001; McCabe et al., 2013; McKenna, 2012). Causes for concern include the growing market power of a handful of private publishers that enables them to set high prices and drain library budgets, socioeconomic gaps in access to research between universities, and an inability to get research into the hands of practitioners, teachers, and policy makers who might use that research to improve lives. This is particularly troubling when you consider that much academic research is either directly funded or indirectly subsidized by public funds and that authors of said research are not directly compensated for their contributions. Indeed, many funding agencies now require that results of funded projects be published *open access* (OA) (Economist, 2012).

Outside the funding agencies, a broader movement is afoot to make OA the new norm in academic publishing. OA means that no pay wall, registration, fee, or subscription is required to download and read material. OA journals (or articles within journals) generate revenue by charging authors a flat fee. This is in direct contrast to traditional pricing models that generate revenue by charging readers an access or subscription fee. The sustainability of the OA business model requires that authors voluntarily pay to publish their work, which in turn requires that OA publications are higher quality than pay-walled publications (McCabe et al., 2013).

For many authors, academic units, and universities the quality of a publication is viewed in terms of page views, downloads, or citations. However, measuring the impact of OA, which is essentially a price change, on a publication's readership is a thorny methodological problem (McCabe & Snyder, 2014). The reason is that there might be various kinds of selection bias associated with the decision to publish an OA article. For example, pay-walled journals might

make their best papers OA to increase the journal's standing, which depends on citation data. Similarly, authors might be more willing to pay OA fees for articles they perceive to be of higher quality. Alternatively, authors might relegate their lower quality work to OA journals if they perceive their chances of publication in those journals to be greater.

Because of these potential selection problems, credible evidence on readers' responsiveness to the cost of academic articles is sparse, and indeed nonexistent in the case of education research. The extant literature is focused on the hard sciences, which were early entrants into OA publishing. For example, McCabe and Snyder (2014) use a journal-volume fixed effects strategy to show that in a sample of top ecology, botany, and multidisciplinary science and biology journals OA increased citations by about 8%. We know of no such analyses in the social sciences or education. We contribute to this gap by providing arguably causal evidence on readers' responsiveness to price, or price elasticity of demand, by exploiting a natural experiment in which several leading education journals unexpectedly became free.

Specifically, in the first two months of 2017, the paywall of the six leading journals of the American Educational Research Association (AERA) went down, essentially granting free access to anyone with a working internet connection. Meanwhile, access to articles in the always-OA *AERA Open* journal was unaffected by this change, as the articles in this journal are always ungated (never behind a paywall). Thus, *AERA Open* serves as a control group in the natural experiment created by the paywall failure. A difference-in-differences style identification strategy can therefore uncover the causal effect of OA on readership by comparing the various journals' downloads before, during, and after the paywall failure.

Using monthly data on downloads, we estimate an event-study time series regression for each of the seven journals. Monthly downloads in the six pay-walled journals increased by 60 to 80% during the two months the paywall was down. This was not due to any confounding

aggregate shock to demand for educational research: downloads of *AERA Open* articles fell by about 45% during this time, suggesting that newfound free access to the other six journals crowded out demand for *AERA Open* articles. However, overall, total journal downloads increased during this time, suggesting it was not a 1:1 crowding out.

Institutional Details and Data

AERA Journals

The American Educational Research Association (AERA) is the largest organization of educational researchers in the United States, with more than 20,000 members. The association operates seven journals, which are described in Table 1. Six of the journals were founded between 1931 and 1979. *Educational Researcher (ER)* and *American Educational Research Journal (AERJ)* publish primarily original empirical research on any aspect of education, with the main difference being the format (*ER* articles are about half the length) and frequency (*ER* is published nine times a year, *AERJ* six). *Review of Education Research (RER)* and *Review of Research in Education (RRE)* are review journals; the distinction is that *RRE* publishes one themed issue per year, whereas *RER* publishes review articles on any topic spread over six issues per year. *Educational Evaluation and Policy Analysis (EEPA)* and *Journal of Educational and Behavioral Statistics (JEBS)* are more specialized journals; the former publishes analyses of the implementation and effects of educational policies (in four issues per year), while the latter publishes research that advances quantitative methods in educational measurement and statistics (in six issues per year). All AERA members are entitled to receive paper copies of *ER* and are allowed to select one of the other five to receive for free; additional journals are available to AERA members for a small fee and to the general public for an annual subscription fee. These six journals are all published by SAGE and are available on SAGE's website behind a paywall.

The newest AERA journal, published since 2015, is *AERA Open*. Like *AERJ* and *ER*, *AERA Open* is a general-interest journal publishing mainly original empirical research on any topic in education. Unlike the other six AERA journals, *AERA Open* is open access. Anyone can go to the SAGE website and download *AERA Open* articles, and there are no paper issues created or disseminated. Articles are generally published on a rolling basis as they are accepted, though they are sometimes rolled out in groups as part of "special topics."

The Natural Experiment

In early 2017, the publisher of the AERA journals, SAGE, switched to an online platform provided by Atypon. To ensure that all subscribers had uninterrupted access during the platform change, the paywall was removed. Ultimately, the platform change was successful and the paywall was down for most of January and February 2017. The timing and duration of the paywall outage were neither announced nor expected, so there were unlikely any strategic or anticipatory responses on the part of readers.

The Data

We have monthly download data from the SAGE website, provided by SAGE and AERA, for each of the seven journals spanning January 2016 to February 2018. Figure 1 displays the raw download data for each journal. The top panel of Figure 1 shows the monthly downloads of each journal, over time, with the ungated period offset by vertical bars. The average journal has about 20,000 full-text downloads per month. Two other patterns are noteworthy. First, the journals themselves vary in download activity, as would be expected given the differences in prestige and topics identified in Table 1. AERJ, ER, and RER tend to have more downloads. Second, a

seasonal pattern in downloads is apparent for all journals: downloads spike early and late in the calendar year.

The top panel of figure 1 also provides suggestive evidence of a causal effect of OA on downloads, as there is a spike in download activity in the six usually-gated journals during the first two months of 2017, but no such spike in *AERA Open* downloads. A similar pattern is observed when averaging the six gated journals together. The bottom panel of figure 1 shows additional evidence of a causal effect by differencing out the same-month, prior-year's downloads. We do this because there is a seasonality to the demand for education research driven by the academic year of university professors and students, who are major consumers of peer-reviewed research, as well as the nine-month school year experienced by many teachers and principals. Again, we see a departure from trend for the six usually-gated journals in January and February of 2017, when the paywall was down. The next section formalizes this analysis using time-series event-study regression models.

Methods

We estimate simple event-study time-series models for each journal (Wooldridge, 2018). The unexpected paywall removal is a shock that potentially disrupts general trends and seasonal patterns in monthly article downloads. Specifically, we estimate models of the form:

$$downloads_t = \theta month_t + \gamma t + \beta ungated_t + u_t, \qquad (1)$$

where *t* indexes the 26 months that span our data, *downloads* is the count of monthly downloads, *month* is a set of month indicators, *t* is a linear time trend, *ungated* is a binary indicator equal to one in January and February of 2017 (when the paywall was down), and zero otherwise, and *u* is a possibly heteroskedastic and serially correlated error term. The parameter of interest is β, which represents the impact of the paywall removal on monthly article downloads.

We condition on linear time trends, though results are robust to using exponential, quadratic, or cubic trends and to ignoring the trend altogether. This is unsurprising, as Figure 1 shows that the trends are relatively flat. More importantly, we condition on a full set of month indicators (fixed effects) to explicitly control for seasonality in downloads. Seasonality is apparent in Figure 1, where there are spikes in download activity in at least some of the journals during March-May and October-November, perhaps in tune with university academic calendars.

We report HAC (heteroskedasticity and auto-correlation) robust standard errors (Newey & West, 1987) that allow for arbitrary forms of heteroskedasticity and auto-correlation (up to 3 lags). We choose three lags because this is the integer suggested by the "quartic-root of T" rule of thumb described in Wooldridge (2018), though the main results are robust to using either 1, 2, or 3 lags and to ignoring possible serial correlation altogether.

We consider two modifications of the model presented in equation (1). First, we replace the *ungated* indicator with two mutually exclusive "January 2017" and "February 2017" indicators, to allow the "no-paywall effect" to vary between the two months the paywall was down. Second, we replace the dependent variable *downloads* with its natural log, which given the non-negative nature of downloads, implements an exponential trend and provides an intuitive semi-elasticity interpretation of β (Wooldridge 2018). Specifically, in the log-level specification of equation (1), $100 \times \beta$ represents the percent change in *downloads* during the *ungated* period (e.g., a coefficient estimate of 0.5 implies a 50% increase in downloads).

We also use estimates of β to compute own- and cross-price elasticities of demand for article downloads. Elasticity is a fundamental concept in economics used to measure the responsiveness of quantity demanded to a price change (Frank, 2014). Elasticities rely on percentage changes in price and quantity, to facilitate comparisons across goods with different

baseline levels of demand. This is appealing given the differences across journals in average monthly downloads.

Of primary interest is the own-price elasticity of demand, which is simply the ratio of the percentage change in quantity demanded (downloads) to the accompanying percentage change in price. When this ratio is less (greater) than one in absolute value, elasticity is said to be inelastic (elastic). Own-price elasticities are commonly computed as $\frac{\Delta Q}{\Delta P} \times \frac{P_0}{Q_0}$, where P_0 and Q_0 represent the initial (pre-change) price and quantity. The change in downloads is simply the estimate of β, the marginal effect of removing the paywall. The change in price is \$36, as the price went from \$36 (P_0) to \$0 for the affected journals. A limitation of this approach is that the estimated elasticity is sensitive to initial price, as it is perhaps arbitrary to distinguish between a price increase of \$36 and a price decrease of \$36 (Frank, 2014). Accordingly, we prefer the arc elasticity, computed via the midpoint formula: $\frac{\Delta Q}{\Delta P} \times \frac{0.5(P_0+P_1)}{0.5(Q_0+Q_1)}$. Here, $0.5(P_0 + P_1) = 18$.

Finally, we compute the cross-price elasticity of demand for *AERA Open*, with respect to the price changes observed in the other six AERA journals. Cross-price elasticities measure the percentage change in demand for good *j* attributable to a 1% change in the price of good *k* (Frank, 2014). The sign of the cross-price elasticity determines whether goods *j* and *k* are complements or substitutes. The preceding discussion of computing and interpreting own-price elasticities applies analogously to the case of cross-price elasticities.

Results

Estimates of various specifications of equation (1) are reported in Table 2. Panel A reports the baseline estimates, where the outcome is total monthly downloads (in 1,000s). Column 1 reports estimates for the *AERA Open* time series, which was OA (freely downloadable) for the entire duration of the time period covered by the analytic sample. The negative point estimate of -4.44

means that when the other six AERA journals' paywalls went down (and became freely downloadable), AERA Open article downloads decreased by about 4,440 per month. Relative to the same months (January and February) of the previous year, when the other AERA journals were paywalled, this amounts to a monthly reduction of about 1/3.

The negative effect on *AERA Open* downloads suggests that there is limited demand for education research articles, and the availability of other free downloads crowded out some potential *AERA Open* downloads. Indeed, the implied cross-price elasticity of -0.2 to -0.4 indicates that *AERA Open* articles are substitutes, not complements, to articles in the other six AERA journals. Still, they are not perfect substitutes and the cross-price demand is relatively inelastic. Moreover, as we see in the subsequent columns of Table 1, the increase in downloads of the other six journals far outweighs the decrease in *AERA Open* downloads, indicating that total downloads went up during the two-month period in which the paywall was down.

Columns 2-7 of Table 1 report the same estimates of the time series specified in equation (1) for each of the six gated AERA journals that experienced a two-month paywall elimination. This effectively reduced the per-download price from $36 to $0. We now explain and interpret the results for the *American Educational Research Journal* (AERJ) in column 2, which is arguably AERA's flagship research journal; estimates for the other journals can be interpreted similarly. The point estimate of 20.92 is positive and strongly statistically significant, suggesting that the elimination of the paywall increased monthly downloads by about 21,000. This is a relatively large effect, representing a 117% increase (more than doubling) over the total downloads over the same two-month stretch in the previous year. This was driven by an effective price reduction from $36 to $0, implying a price elasticity of demand for AERJ articles of 1.17. However, an arguably more useful estimate of the elasticity is the arc elasticity, which is computed at the midpoint of the price and quantity changes, and is thus independent of whether

the starting price was $36 or $0. The implied arc elasticity of demand for AERJ article downloads is 0.37, which is significantly smaller and implies inelastic demand for AERJ downloads.

The results for AERJ are quite similar to those for the other five usually gated AERA journals, as reported in columns 3-7 of Table 1. While the point estimates vary in magnitude from effects of about 3,000 downloads for the *Journal of Educational and Behavioral Statistics* (JEBS) to more than 38,000 for the *Review of Educational Research* (RER), all are positive and strongly statistically significant. Interestingly, the variation in absolute effect sizes across journals is largely driven by the underlying average monthly download rates of those journals. Expressed as percentage increases from downloads in the same months of the previous year, the effects map into a range of 80% to 150% increases and arc elasticities ranging from 0.29 to 0.42. It is interesting and intuitive that the smallest effects and elasticities are observed for JEBS while the largest effects and elasticities are observed for RER, as JEBS is arguably the most technical and specialized of the six journals, while RER has the largest impact factor (and AERJ, the owner of the second largest effects and elasticities) intentionally target a broader audience.

Panel B explores the dynamics of the temporary paywall removal, which lasted two months, by differentiating the effect in month 1 (January 2017) from that in month 2 (February 2017). In all cases, the two month-specific effects are strongly jointly significant. Column 1 shows that the crowd-out effect of ungated access to the other AERA journals on the open-access *AERA Open* seen in Panel A occurs primarily in January, as the effect on downloads in February remains negative, but is indistinguishable from zero and significantly smaller than the January effect at the 5% confidence level. This is perhaps partly explained by the significant (50%) drop in *AERA Open* downloads observed between January and February in 2016.

Columns 2-7 of Panel B show that the effect of the paywall removal on downloads of the six previously gated AERA journals was larger in February than in January. These differences are at least marginally significant for each journal except for JEBS, which again might be due to the technical and specialized nature of JEBS. Perhaps some of these monthly differences were due to interested parties learning about the fallen paywall, but more likely it has to do with seasonal patterns in downloads: there are more downloads in February than in January, as seen in the 2016 figures. Indeed, in percentage terms, the monthly effects tend to be within 20 to 30 percentage points of each other, and constitute similar percentage increases over the same month's downloads in 2016 as the baseline results in Panel A (80% to 150%).

Finally, Panel C estimates a version of equation (1) that takes the natural log of monthly downloads as the outcome variable. These are our preferred estimates, as they account for the strictly non-negative nature of monthly downloads, allow for an exponential trend, and provide a useful semi-elasticity interpretation of the paywall effect (Wooldridge 2018). Specifically, the point estimates in Panel C represent semi-elasticities, suggesting that the removal of the paywall causes a 45% reduction in *AERA Open* monthly downloads and, for example, a 71% increase in AERJ monthly downloads. Once again, the point estimates are positive and strongly statistically significant for each of the six previously gated AERA journals and range from effects of 58% (JEBS) to 83% (RER). These results are broadly consistent with those reported in Panel A and provide robust evidence of a sizable, arguably causal effect of OA on journal downloads.

Conclusion

We estimate the causal effect of open access to academic journal articles on article downloads by exploiting a natural experiment in which the paywall to several prestigious educational research journals was unexpectedly taken down for two months. Using the always open-access *AERA*

Open as a control group, we find credibly causal evidence that removing the paywall increased monthly downloads by 60 to 80%. This is a fairly large effect on downloads and suggests that thousands of potential readers interested in educational research are stymied each month by paywalls.

Our focus on downloads is necessitated by data availability, as this is the only proxy for article consumption available to us. Nonetheless, downloads are arguably a useful proxy for an article's impact. We are of course unable to address the important question of how many downloads are actually read, and of those read how many alter the reader's behavior or knowledge. Previous work has focused on citations rather than downloads (e.g., McCabe & Snyder 2014, 2015). We cannot say how many of the downloads induced by the temporary period of OA led to citations, though descriptive research on the determinants of citations find that every 50 to 140 downloads lead to one citation (Gorraiz et al. 2014). Assuming that about 1/100 (1%) of downloads lead to a citation suggests that the temporary removal of the paywall generated anywhere from 30 (JEBS) to 250 (ER) citations per journal, per month. That said, in the case of educational research, citations are not necessarily the correct measure of impact, as teachers, schools, and districts may use journal articles to inform their policy and practice without ever formally citing the research.

Future work should assess the economics of publishing educational research more generally, including who opts in to *AERA Open*, who pays the submission fees, perceptions and citation rates of the quality of *AERA Open* and other OA journals relative to subscription-based journals, and the price sensitivity of university libraries to journal subscription prices.

REFERENCES

Bergstrom, T. C. (2001). Free labour for costly journals? *Journal of Economic Perspectives*, *15*(4), 183-198.

Economist. (2012). Brought to book: Academic journals face a radical shake-up. The Economist. July 21, 2012. Accessed online January 29, 2019.

Frank, R. H. (2014). *Microeconomics and Behavior*. New York: McGraw Hill.

Gorraiz, J., Gumpenberger, C., & Schlögl, C. (2014). Usage versus citation behaviours in four subject areas. *Scientometrics*, *101*(2), 1077-1095.

McCabe, M. J., & Snyder, C. M. (2014). Identifying the effect of open access on citations using a panel of science journals. *Economic Inquiry*, *52*(4), 1284-1300.

McCabe, M. J., & Snyder, C. M. (2015). Does online availability increase citations? Theory and evidence from a panel of economics and business journals. *Review of Economics and Statistics*, *97*(1), 144-165.

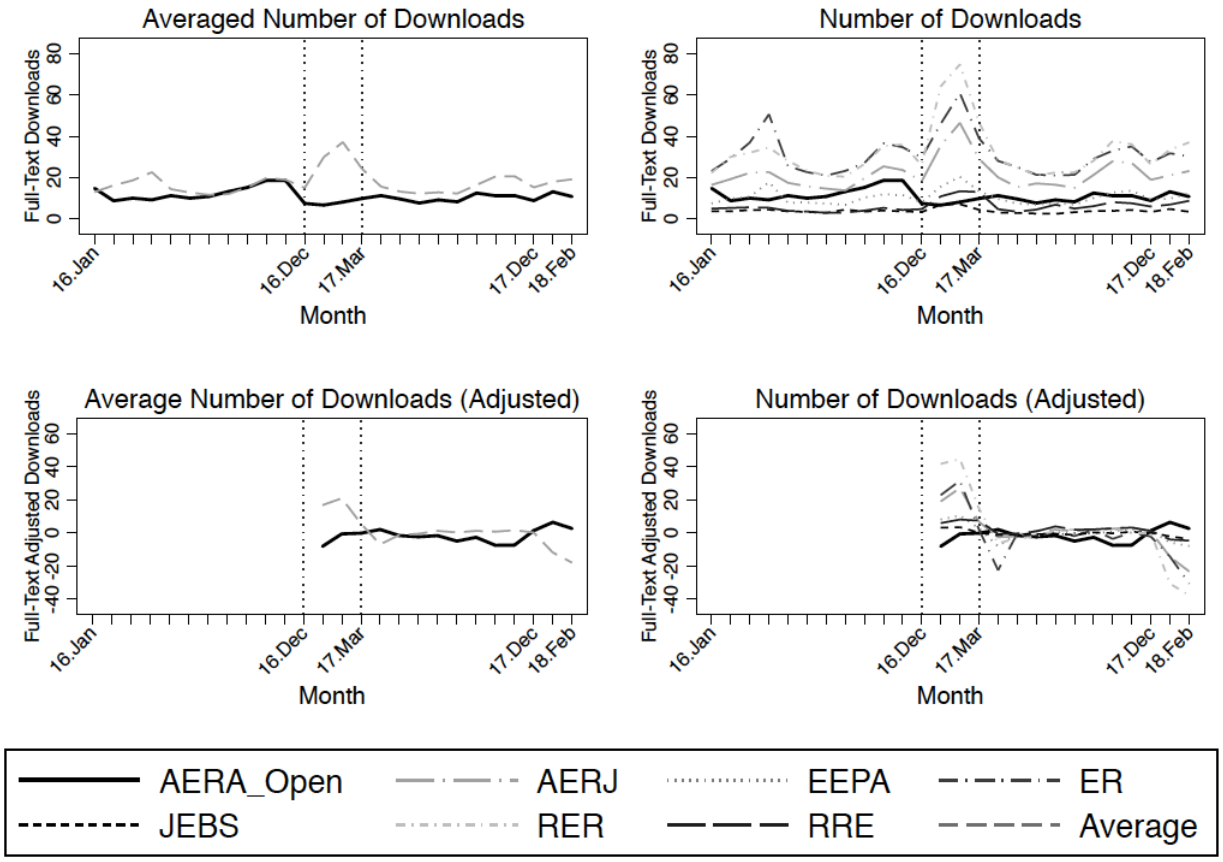
McCabe, M. J., Snyder, C. M., & Fagin, A. (2013). Open access versus traditional journal pricing: Using a simple "platform market" model to understand which will win (and which should). *The Journal of Academic Librarianship*, *39*(1), 11-19.

McKenna, L. (2012). Locked in the Ivory Tower: Why JSTOR imprisons academic research. The Atlantic. January 20, 2012. Accessed online January 28, 2019.

Newey, W., & West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703-708.

Wooldridge, J., (2018). *Introductory Econometrics*. Mason, OH: Cengage Learning.

Figure 1. Monthly Downloads, in 1,000s



Notes: Figures in the left column report monthly downloads for AERA Open, which is always ungated, and the monthly average of the other six journals that are normally gated, but were ungated in January and February 2017. The figures in the bottom row are "adjusted" by differencing out the same month's download total in the previous year.

Table 1: Seven AERA Journals

Journal Title	Impact Factor	Rank in Education Journals	Individual Subscriptions	Institutional Subscriptions	Year Founded
	(1)	(2)	(3)	(4)	(5)
Review of Educational Research	8.24	1/238	6031	9006	1931
American Educational Research Journal	2.46	33/238	11147	8869	1964
Educational Researcher	4.00	6/238	25529	8615	1972
Review of Research in Education	1.59	91/238	n/a	n/a	1973
Journal of Educational and Behavioral Statistics	2.23	45/238	1980	8344	1976
Educational Evaluation and Policy Analysis	2.48	30/238	3276	8448	1979
*AERA Open**	n/a	n/a	n/a	n/a	2015

* *AERA Open* is an open access (OA) journal. It is the only AERA journal that is OA. It is also fully online. One-year impact factors and rankings are from 2017 inCites Journal Citation Reports published by Web of Science.

Table 2: Time Series Event Study Estimates of Effect of Taking down Paywall on Monthly Downloads

Journal	AERA Open	American Educational Research Journal	Educational Evaluation and Policy Analysis	Educational Researcher	Journal of Educational and Behavioral Statistics	Review of Educational Research	Review of Research in Education
Acronym	AERA Open	AERJ	EEPA	ER	JEBS	RER	RRE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Monthly Downloads, in 1,000s							
No Paywall	-4.44**	20.92***	8.07***	24.75***	3.08***	38.61***	5.65***
	(1.77)	(1.60)	(0.90)	(3.66)	(0.37)	(2.10)	(0.57)
E[Y|t = 1, 2]	11.86	17.92	8.75	26.25	3.81	26.25	5.25
Implied ε|p=36	-0.37	1.17	0.92	0.94	0.81	1.47	1.08
Implied arc ε	-0.23	0.37	0.32	0.32	0.29	0.42	0.35
B. Monthly Downloads, in 1,000s, month-specific effects							
No Paywall 1	-7.28***	16.69***	6.95***	18.53***	2.62***	36.13***	4.95***
	(0.55)	(0.59)	(0.82)	(5.15)	(0.67)	(3.19)	(0.94)
No Paywall 2	-1.61	25.15***	9.20***	30.97***	3.54***	41.09***	6.36***
	(2.56)	(0.19)	(0.97)	(1.03)	(0.01)	(0.95)	(0.14)
E[Y|t = 1]	14.88	16.51	7.61	23.22	3.87	22.44	5.08
E[Y|t = 2]	8.84	19.33	9.89	29.27	3.75	30.06	5.41
Joint Sig. (p)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
H_0(1 = 2) (p)	0.03	0.00	0.01	0.02	0.19	0.08	0.12
C. Logged Monthly Downloads							
No Paywall	-0.45**	0.71***	0.60***	0.62***	0.58***	0.83***	0.64***
	(0.16)	(0.04)	(0.08)	(0.10)	(0.09)	(0.07)	(0.09)

Notes: T = 26 monthly data points, from January 2016 (t = 1) through February 2018 (t = 26). The "No Paywall" period lasted two months: January (1) and February (2) of 2017. All models condition on month fixed effects and a linear time trend. The results are robust to using a quadratic time trend or eliminating the trend altogether. HAC (heteroskedasticity and autocorrelation) robust standard errors are reported in parentheses. Specifically, these are Newey-West standard errors that allow for three lags, which was determined following the "quartic-root of T" rule of thumb. However, the results are robust to allowing for 1, 2, or 3 lags. For each model, we report the mean of the outcome for the analogous pre-treatment month(s). E.g., panel A reports mean downloads in January and February of 2016 to provide a reference point for interpreting the effect of the January-February 2017 treatment effect. Elasticities (ε) are computed based on the stated per-article download price of $36 becoming $0 during the treatment period of no paywall. Column 1 reports cross-price elasticities; the negative sign implies that AERA Open articles are substitutes, not complements, to the other journals' articles. ***, **, *, indicate statistical significance at the 0.01, 0.05, and 0.1 levels, respectively.