

The Play's the Thing: Experimentally Examining the Social and Cognitive Effects of School Field Trips to Live Theater Performances

Jay P. Greene¹, Heidi H. Erickson¹, Angela R. Watson¹, and Molly I. Beck¹

Field trips to see theater performances are a long-standing educational practice; however, there is little systematic evidence demonstrating educational benefits. This article describes the results of five random assignment experiments spanning 2 years where school groups were assigned by lottery to attend a live theater performance or, for some groups, watch a movie version of the same story. We find significant educational benefits from seeing live theater, including higher levels of tolerance, social perspective taking, and stronger command of the plot and vocabulary of those plays. Students randomly assigned to watch a movie did not experience these benefits. Our findings also suggest that theater field trips may cultivate the desire among students to frequent the theater in the future.

Keywords: arts education; educational policy; educational reform; experimental design; instructional practices; museum education; student behavior/attitude

Hedberg, 2011). Cultural field trips face severe cutbacks if they cannot show improved performance on tested subjects or other important educational benefits.

This article describes the results of a series of five experiments conducted over 2 years in which school groups from a variety of grades were assigned by lottery to see live theater. We find significant educational benefits from taking students to a theater performance. In particular, students in our full sample randomly assigned to see live theater demonstrate higher levels of tolerance and social perspective taking (SPT) as well as stronger command of the plot and vocabulary of those plays. When results are broken out for each of the five plays in our study, point estimates for each of these outcomes remain positive but sometimes fall short

of statistical significance due to smaller sample sizes. Our findings also suggest that theater field trips may cultivate the desire among students to frequent the theater in the future, although this result is only marginally significant.

In addition, for two of the five experiments, a second treatment condition was added in which some students were randomly assigned to see a movie comparable to the play seen by other groups of students. Leaving school to see a movie did not produce the same benefits as viewing live theater. The evidence suggests that there are educational benefits to the traditional practice of school field trips to see plays and that those benefits are unlikely to be replicated by showing students movies instead.

¹University of Arkansas, Fayetteville,

AR

Student field trips to see live theater performances are

^a long-standing educational practice. Like many common school practices, however, there is little systematic evidence demonstrating educational benefits. Field trips to see plays continue mostly with the support of the wisdom of educators and a sensible deference to prior practice. With the rise of test-based accountability, however, many traditional school practices are under pressure, including school field trips to theaters and other cultural institutions (Gadsden, 2008; Rabkin &

Previous Research

Attending arts-focused school field trips is a longstanding tradition in American public education; however, there is mounting evidence that students are receiving less exposure to the arts, both from in-school and out-of-school activities. Cultural institutions in large cities, such as Chicago and Cincinnati,

report a marked drop in visits from school groups (Greene, Government Accountability Office (GAO) report on access to Kisida, & Bowen, 2014). A survey administered by the arts education found 7% of public school teachers reported a decline in the amount of instructional time dedicated to arts that schools were increasingly cancelling previously planned education in schools—a modest decrease (GAO, 2009). Of field trips (Ellerson & McCord, 2009). In addition, more concern, the GAO found that schools with high numbers

of minority students and greater academic challenges were (Møllegarrd, 2017, p. 142). The researchers then examined how significantly more likely to report a decline in arts instructional well each child did according to time, including arts-focused field trips (GAO, 2009). Further, principals reported struggles with decreased budgets and demands placed on instructional time as a result of accountability standards and voiced concerns that arts education might suffer cuts (GAO, 2009). The GAO recommended further study of these issues to ensure that all students receive a well-rounded education. Similarly, the National Endowment for the Arts (NEA) reports a decline in adult attendance and participation in the arts and that the greatest predictor of future arts consumption and participation is arts education (Rabkin & Hedberg, 2011).

While there has been little rigorous research that speaks directly to the effects of seeing live theater on students, there is a growing literature on related topics. For example, a recent, large-scale experiment found that a single school field trip to tour an art museum caused significant effects that could be observed nearly 2 months following the visit (Greene et al., 2014). Students randomly assigned to receive the art museum tour were significantly more likely to be interested in visiting cultural institutions in the future and actually did so at a higher rate than students randomly assigned to the control group that had not toured the museum (Kisida, Greene, & Bowen, 2014). Students who toured the art museum also scored significantly higher on a measure of their ability to think critically about works of art (Bowen, Greene, & Kisida, 2014). In addition, students who visited the art museum displayed higher levels of content knowledge, tolerance, and historical empathy as a result of their tour (Greene et al., 2014). All of these benefits were more likely to be realized by students from more disadvantaged backgrounds, suggesting that advantaged parents can more easily substitute with their private efforts if schools fail to take students to cultural institutions. It appears that schools may play an essential role in providing equal access to cultural institutions and any benefits they produce.

In addition, some long-term analyses find that exposure to cultural activities improves academic outcomes for students years afterwards. For example, Jægar and Møllegarrd (2017) studied a large sample of monozygotic twins in Denmark to see if their cultural activity was related to later educational outcomes. By comparing outcomes within pairs of identical twins, the researchers control automatically for a large set of unobserved environmental and genetic factors. The mothers of these sets of identical twins were asked about each child's cultural activity at the age of 12, including "how often child went to any type of museum" and "how often child went to the theater or a musical performance" (Jægar &

teacher and independently proctored grades at the age of 15 as well as the rate at which they graduated from high school. There were strong correlations between the relative level of cultural activity in which each twin engaged, their independently proctored grades, and their likelihood of graduating from high school. Other long-term studies find correlations between student involvement in the arts and later academic outcomes (Lacoe, Painter, & Williams, 2016; Ruppert, 2006), but most of these studies have difficulty establishing that this is a causal relationship.

Goldstein and Winner (2012) conducted a set of experiments to examine how students are affected by drama activities and find significant social-emotional benefits measured shortly after the intervention. There is also an extensive literature that examines how drama-based instructional techniques affect students. In a meta-analysis of that research, Lee, Patall, Cawthon, and Steingut (2015) found that drama-based pedagogy can have significant academic effects as well as social-emotional outcomes. While this is a comprehensive review of that research, the authors acknowledged that the 47 studies they examined are quasi-experimental, not experimental, and therefore some of the observed relationships may not be causal. In another meta-analysis of the literature examining the value of arts integration, the American Institutes for Research (AIR) finds that the average child could gain up to four percentage points in achievement from arts integration interventions, but almost none of the studies used to generate this estimate utilized the strongest causal research designs (Ludwig, Boyle, & Lindsay, 2017).

We have some evidence to suggest that students benefit from school visits to art museums, experience long-term academic gains from frequenting museums and the theater, and may learn from drama-based pedagogy and theater activities. But regarding the exact question addressed in this study—whether students benefit from school visits to see live theater—there is little direct evidence. The results of two of the five theater experiments contained in this article were described in an earlier publication (Greene, Hitt, Kraybill, & Bogulski, 2015), but to our knowledge, this is the first large-scale experiment to examine what students learn from seeing live theater.

Research Design

This study addresses the question of whether students benefit

from school visits to see live theater. School groups were randomly chosen to receive free tickets to attend one of five theater performances over a 2-year period. Participating schools were primarily from the Northwest corner of Arkansas. Most schools were in semirural areas with over half of students qualifying for free or reduced priced lunch. Teachers applied for a chance to bring their class to one of five different theater performances. Teachers and students were unaware of the hypotheses being tested at any time during the experiment. We then matched applicant classes based on their similarity to student populations. Typically, applicant classes were matched with others within their same school, but sometimes they were matched with classes in other schools that had similar students. Matched students within the same or neighboring schools were likely to share similar observed and unobserved characteristics. Within

the matched set, we randomly assigned one or more classes to receive tickets and one to serve as a control group, ensuring that treatment and control groups were similar. While treatment groups received a field trip, the control groups did not and continued with their normal school activities. In total, we conducted 47 lotteries, creating 94 treatment and control groups containing almost 1,500 students. Performances included *A Christmas Carol*, *Hamlet*, *Around the World in 80 Days*, and *Peter and the Starcatcher*, all performed by an award-winning professional company, and *Twelfth Night*, performed by university theater students.

For the final two plays, *Peter and the Starcatcher* and *Twelfth Night*, we were able to add a second treatment condition in which students would be randomly assigned to see a movie that was similar to what the theater treatment group saw. For the *Twelfth Night* experiment, applicant groups were randomly assigned to see the play, to see the 1996 film of the same Shakespeare story, or to serve in the control group, which saw neither the play nor the movie. While the mode of delivery was different, and indeed that is the point, the play and the movie both used virtually the same script for the performances. For the *Peter and the Starcatcher* experiment, applicant groups were randomly assigned to see the play, to see the 1991 film *Hook*, or to serve in the control group that saw neither. In this case, the theater performance and the movie used different scripts but had similar content, characters, and themes. This addition of a movie treatment allowed us to test whether any effects of seeing a play were derived from the subject matter of the play or from the experience of seeing live theater. The fact that both play and movie treatment groups left school on the same bus, at the same time, and only differed in whether they

walked into a ballroom to see the movie or into the theater to see the play allows us to examine whether any observed outcomes could be caused simply by leaving school on a field trip. While the play and movie may not have been identical in both cases, the policy-relevant question we address is whether students are differentially affected by the two modes of delivery. If watching a movie is as effective as taking a bus to see a play, then students might simply watch movies. But if the play is more effective than a virtually identical or similar movie, then the allocation of resources in both time and money would be justified for field trips to see live theater.

It is important to note that the experimental interventions did not include anything beyond the opportunity to see live theater. The treatment did not provide any additional training, materials for teachers, or supplemental activities for students. Any supplementary activities, such as reading the play or watching the movie in class, were assigned at the teachers' discretion and could have occurred among treatment or control groups. If teachers in the treatment group were more likely to provide additional instruction or discussion, then that is part of the treatment effect we are estimating. Our lack of information on the extent to which teachers may have prepared their students to see the plays, however, prevents us from knowing with confidence the extent to which outcomes are attributable to the play alone or to any preparation.

To collect outcome measures, we administered surveys to treatment and control students in their classrooms. On average, surveys were administered 54 days after the treatment group had

248 EDUCATIONAL RESEARCHER

seen the play. There were not differential participation rates among the treatment and control groups. We collected surveys from 77.6% of the students assigned to see a play, 76.0% for those assigned to see a movie, and 76.5% among control group students.

While the basic design—offering free tickets, matching similar applicant groups, and then conducting a lottery within matched sets of applicants—remained the same across all five plays, some important details did change over time. For example, after the first two plays, we changed the survey to add a scale designed to measure students' SPT (Gehlbach, 2004; Gehlbach et al., 2008) while dropping the Reading the Mind in the Eyes Test (Baron-Cohen, Wheelwright, Hill, & Raste, 2001). We believed that SPT would be a better way of capturing potential social-emotional effects of seeing live theater.

In addition to replacing one measure on the surveysuspected that by exposing students to a broader world through instrument after the first two plays, we also administeredtheater, they would increase their ability to understand other pretreatment surveys to students after the first three plays. Inpeople’s points of view (SPT) as well as gain greater prior administration, we lacked the resources to collectacceptance of other people (Tolerance) (Gehlbach, 2004; measures both before and after the intervention, so we relied onGreene et al., 2014). We also sus- pected, based on prior work the lottery to give us equivalent treat- ment and control groupsby Greene et al. (2014), that seeing live theater would be an since we could not control for any pre- treatment differences.effective mechanism for conveying the plot and vocabulary of For the final two plays, however, we were able to administerthese plays, so we included measures of Content Knowledge. surveys to all students both before and after the treatmentIn addition, since past research suggested that visiting cultural occurred. This allowed us to check whether our treat- ment andinstitutions increased the desire to frequent those institutions in control groups were similar on pretreatment measures of thethe future (Kisida et al., 2014), we included measures of outcomes. Controlling for pretreatment measures of theTheater Consumption in the survey. Lastly, we included outcomes also improves the precision of our estimates of treat-measures of the desire to Participate in Theater because we ment effects, although it reduces the size of the sample, whichthought it was possible that seeing live theater might inspire reduces the precision of estimates. On balance, we prefer thestudents to become more involved in theater by auditioning for mod- els using the larger sample and without controlling forplays and taking drama class. baseline measures because the larger sample tends to give us smaller stan- dard errors with generally the same point estimates as the smaller samples that control for baseline measures.

Implementation of Research Design

Other outcomes measured in the survey remainedcontrol groups confirms that randomization was successful in unchanged across the five plays. Given prior research that hadhelp- ing ensure that we compared generally similar groups. found social- emotional benefits from arts-focused field trips,There were we used a scale to measure Tolerance (Greene et al., 2014). We

Examining the background characteristics of our treatment and

Table 1 Student Characteristics in Treatment and Control Groups

Control Group Mean Play Treatment Mean Movie Treatment Mean *N*

Full sample

Female 0.609 0.550* 0.596 1,467 Age 14.751 14.717 15.175 1,463 Grade 8.628 8.605 9.059 1,485 White 0.672 0.660 0.681 1,467 Seen play 0.304* 1,467

Twelfth Night and *Peter and the Starcatcher* sample with pretest

Tolerance 2.986 2.936 2.959 485 Social Perspective Taking 2.557 2.451 2.471 485 Content Knowledge 0.368 0.394 0.354 485 Theater Consum 485 Theater Participation 1.961 1.676 1.870 485

Note. The outcomes of interest (Tolerance, Social Perspective Taking, Comprehension, Vocabulary, Theater Consumption, and Theater Participation) were only prete to see *Twelfth Night* and *Peter and the Starcatcher*. * $p < .10$. ** $p < .05$. *** $p < .01$ (two-tailed, relative to control group).

no differences in background characteristics that were significant at $p < .05$ (see Table 1). There were two instances in which differences are significant at $p < .10$, but we might expect this by chance given that we were comparing three groups on 10 different variables. On our pretreatment measures of outcomes, we observed no statistically significant differences between the control and treatment groups.

grade and some as old as 12th grade. Approximately two thirds of our sample identified as white, which reflects the broader community in which the experi- ments took place. About one quarter of students had seen a play in the previous year, however this might have included school plays, church plays, and holiday shows. This probably reflects a relatively low level of previous exposure to theater and cultural activity.

Students in our sample were just shy of their 15th birthday and in the middle of 9th grade on average. However, students varied in age, with some students as young as 4th

Two problems occurred during implementation of the research design that caused some applicant groups not to see the play despite being assigned to the play treatment. Severe winter weather forced the cancellation of a performance of *A*

Christmas Carol. Additionally, during the *Around the World in 80 Days* experiment, the theater made an error that caused the actors not to be available to perform when students arrived. To be very conservative, we report results for Intention to Treat (ITT), in which we count all students randomly assigned to the treatment group as if they received the treatment even if they failed to do so because of weather or a scheduling error. To provide a more realistic estimate of the treatment effect, we also report the Impact on Treated (IOT) derived from a two-stage model in which the first stage uses assignment to the treatment group as a predictor of whether students actually received the treatment.

Outcome Measures

The Tolerance scale consisted of seven items. Students were given four options, from *strongly disagree* to *strongly agree*, to respond to a

series of statements designed to capture their general acceptance of other people and different opinions. The scale was adapted from Greene et al. (2014) and included statements such as “People who disagree with my point of view bother me” or “I think people can have different opinions about the same thing.” The Cronbach’s alpha for this scale was .71, suggesting that there is an acceptable amount of internal consistency within this scale.

The SPT scale also consisted of seven items and was adapted from Gehlbach, Brinkworth, and Wang (2012). Students had five response options, from *almost never* to *almost all of the time*, to questions like “How often do you try

The Theater Consumption scale consisted of 10 items adapted from Kisida et al. (2014). Students had four response options to interested are you in seeing live performances in a theater?” or statements like “I plan to see live theater performances when

The Theater Participation scale was also adapted from Greene et al. (2014) and consisted of four items. Students were asked interested are you in taking a drama class?” and “If your school were having auditions for a new play, how interested would you role in that play?” The Cronbach’s alpha for this scale was .93.

In Table 1 all of these outcomes are expressed as the mean of a 0–3 or 0–4 scale, except for Content Knowledge, which is expressed questions answered correctly. For the purposes of the outcome analyses, all scales were converted into *z* scores with a mean deviation of 1. The reported results, therefore, are the effect sizes expressed as a percentage of a standard deviation.

Analyses

Because the randomized controlled trial research design used here has the important feature of generating comparable treatment (which we confirmed with the analyses presented in Table 1), we can use a straightforward set of analytic techniques, designed experiments, to estimate the impact of a school field trip to see live theater on student outcomes. In its simplest form, this technique differences using the following equation for outcome Y of student i in matched set m :

$$Y_{im} - \alpha - \beta_1 \text{Play}_{im} - \beta_2 \text{Movie}_{im} - \beta_3 \text{Match}_{im} = \epsilon_{im} \quad (1)$$

where the binary variable Play_{im} is equal to 1 if the student is in the treatment group that was randomly assigned to receive free tickets to see one of the five plays and is equal to

to figure out what motivates others to behave as they do?” and “Overall, how often do you try to understand the point of view of other people?” The Cronbach’s alpha in our study for this scale was .85, indicating strong internal consistency in students’ responses.

Our measure of Content Knowledge consisted of six questions about the plot of each play and five questions about vocabulary drawn from the play. For example, for students in the *Hamlet* experiment, we asked, “What happens to Ophelia?” or asked them about the definition of “countenance” and provided four response options to each question. It is reasonable to expect that students who attended the play or movie would be more likely to know the plot and vocabulary than students who saw neither. However, it is not guaranteed that simply taking students to the theater means that they will acquire knowledge about what they saw. We are able to determine if students actually learn and retain content knowledge from attending a play or watching a movie by comparing treatment and control students. In addition, it is important to measure the content knowledge of control group students because students may acquire information about the plot and vocabulary of these shows from other sources without seeing a play or movie. Given that the questions were different for each play, calculating a combined Cronbach’s alpha is not possible, but we are confident that we captured meaningful variation in knowledge about each play’s plot and vocabulary.

Table 2 Impacts of Play and Movie Treatment on Five Primary Student Outcomes

Outcome	Play Treatment	Movie Treatment	Female	White	N
Tolerance	0.142**	-0.038	0.290***	0.231***	1,441
Social Perspective	0.190**	-0.009	0.289***	0.235***	1,441
Content	0.169**	-0.045	0.407***	0.090	791
Theater Taking	0.222**	-0.014	0.406***	0.102	1,442
Knowledge	0.101***	-0.015	-0.016*	0.060***	1,442
Consumption	0.135***	0.006	-0.016*	0.063***	1,442
Theater Participation Intention to Treat	0.127*	0.110			
Impact on Treated	0.170*	0.110			
Theater Participation Intention to Treat	0.127*	0.110			
Impact on Treated	0.170*	0.110			
Theater Participation Intention to Treat	0.127*	0.110			
Impact on Treated	0.170*	0.110			
Theater Participation Intention to Treat	0.127*	0.110			
Impact on Treated	0.170*	0.110			

Note. Standard errors—clustered within classes—are presented in parentheses. Coefficients for age and for the matched set parameters are not presented. C baseline were taken for students who saw *Twelfth Night* and *Peter and the Starcatcher* only. * $p < .10$. ** $p < .05$. *** $p < .01$ (two-tailed).

dent otherwise. is in the The treatment binary variable group that *Movie*_{*im*} is randomly equal to 1 assigned if the stu- to t movie and is equal to 0 otherwise. Because the groups were created using a randomization proce- dure the model within applicant binary variables sets, *Match*_{*m*} is have also the effect of estimating within, as opposed to across, matched group- ings. Finally, ϵ_{im} is a stochastic error term clustered at the cl- into account the spatial correlation from students nested within classrooms.

Proper randomization generates experimental groups that are comparable but not necessarily identical. The basic regression mo- improved by adding controls for observable char- acteristics to increase the reliability of the estimated impact by accounting for improving the precision of the overall statistical model. This yields the following equation:

$$Y_{im} = \alpha + \beta_1 Play_{im} + \beta_2 Movie_{im} + \beta_3 Match_m + \beta_4 Gender_{im} + \beta_5 Age_{im} + \beta_6 White_{im} + \epsilon_{im}, \quad (2)$$

where *Gender*_{*female*} and precise age 0_{im} otherwise, is of student a binary variable equal to 1 if the student is a *Age*_{*i*} at the posttreatment variable surveys indicating were administered, student does not and identify *White*_{*im*} as being is a binary white va- and is 1 otherwise. In this model, effect of β_1 class and β_2 field are trip the for parameters students of in interest and represent t treat- ment groups. Equation 2 is our preferred model and was used to produce the ITT results presented in Table 2. The ITT estim- effect was for all students assigned to the treat- ment, regardless of whether they actually attended the field trip. It is the most cc the treatment effect because it could be biased downward due to some students not seeing the play despite being Due to the noncompliance to treatment assignment during the first and third play experiments caused by weather and a scheduli-

the theater, we are also in generating an IOT estimate. The IOT estimate describes what the effect would have been had all of the applicant groups actually they were randomly assigned. The model used to generate that estimate is a two-stage least squares model in which the se to Equation 2 except that ment $Play_{im}$ is derived to predict from treatment the first compliance.

stage in which lottery assign-

Because we do not have strong theoretical expectations that different plays should produce different effects and because the individual play is relatively small, we present in Tables 2 and 3 the results of all five plays combined. The out- comes for SPT in the results for students who saw the last three plays since SPT was not added to the survey instrument until that time. All of the ot 2 represent the results of students across all five plays.

For the last two plays, we surveyed all students prior to the treatment and again after the intervention, which allows us to contri measure of the outcome. For exam- ple, we can control for students' score on the Tolerance scale prior to the intervention when e the treat- ment on their Tolerance score collected after the intervention. The model we used to generate these results (as presen expressed as:

$$Y_{im} = \alpha + \beta_1 Play + \beta_2 Movie + \beta_3 Match + \beta_4 PreTest_{im} + \beta_5 Age_{im} + \beta_6 White_{im} + \epsilon_{im} \quad (3)$$

This is identical to Equation 2 except the pretreatment measure of the that outcome it adds $PreTest_{im}$, including im , which Tolerance, SPT, Content Knowledge, and Theater Consumption and Participation.

Results

As shown in Table 2, providing students with the opportunity to leave school on a field trip to see a live theater performance Table 3 Impacts of Play and Movie Treatment on Five Primary Student Outcomes Controlling for Base

Intention to Treat

Tolerance

Social Perspective

Taking Content Knowledge

Theater Consumption

Theater Participation

Play treatment	0.182**	0.056	0.154***	0.130	-0.040				
	(0.071)	(0.076)	(0.025)	(0.081)	(0.064)	Movie treatment	0.074	-0.064	0.051*
	(0.074)	(0.064)	(0.028)	(0.066)	(0.057)	Pretest	1.046***	0.681***	0.424***
	(0.066)	(0.051)	(0.057)	(0.045)	(0.025)	Female	0.071	0.312***	-0.013
	(0.081)	(0.094)	(0.015)	(0.076)	(0.073)	White	0.052	-0.005	0.040*
	(0.099)	(0.062)	(0.021)	(0.065)	(0.073)	N	475	475	475

Note. Standard errors—clustered within classes—are presented in parentheses. Coefficients for age and for the matched set parameters are not presented. C baseline were taken for students who saw *Twelfth Night* and *Peter and the Starcatcher* only. ** $p < .05$. *** $p < .01$ (two-tailed).

produced a number of significant effects. Students given the opportunity by lottery to see a play scored .142 of a standard d Tolerance scale than if they were in the control group. If we adjusted for the noncompliance produced by bad weather and a estimated effect of actually seeing a play on Tolerance increases to .190 of a standard deviation. Being assigned to see a movie ins to have no effect on Tolerance. When results are broken out for each of the five plays in our study, point estimates for Toleran sometimes fall short of statistical significance due to smaller sample sizes.

SPT increases by .169 of a standard deviation for students randomly assigned to go on a school field trip to see a play. Again, those who actually saw the play. Being assigned to see a movie, however, has no effect on SPT. Student's Content Know vocabulary in these stories is also increased when students see the play. Watching a movie did not convey this Content Knowl seeing the live performance. When results are bro- ken out for each of the three plays in our study for which we measured SP estimates remain positive, but none of them are statistically significant due to smaller sample sizes. Taking a field trip to see a

student interest in consuming theater in the future but that effect was only significant at $p < .10$. Being assigned to the movie similar estimated effect, but it was not statistically significant even at $p < .10$. Neither the play nor movie treatments had an student interest in participating in future theater activities.

Results for each play are presented in Table 4; the point estimates are roughly consistent across all plays. That is, if the overall statistically significant, the individual play estimates are almost all positive and many are also statistically significant. The only the analyses of each individual play has to do with the effect of the movie

MAY 2018 251

treatment on Content Knowledge. Seeing a movie of *Twelfth Night* increased understanding of the plot and vocabulary of the control group by .045 of a standard deviation but that effect is only statistically significant at $p < .10$ and is almost one the treatment effect. Seeing the movie *Hook* had no effect on the Content Knowledge for *Peter and the Starcatcher* but that is not to

the stories are not as well aligned as are the movie and play of *Twelfth Night*. When we control for pretreatment measure presented in Table 3, the results generally remain the same. The Tolerance and Content Knowledge effects from the p approximately the same size and continue to be statistically significant. The result for SPT shrinks in

252 EDUCATIONAL RESEARCHER

Table 4 Intention to Treat Impact on Each Outcome of Interest for Each Play Combined Plays

*A Christmas
Around the
Twelfth Carol Hamlet
World in 80 Days
Night*

Peter and the Starcatcher

Impact on Tolerance

Play treatment 0.142** 0.231 0.261** 0.288** 0.337*** 0.076
(0.065) (0.145) (0.118) (0.100) (0.110) (0.176) Movie treatment -0.038 0.144 -0.126
(0.095) (0.124) (0.193) N 1,441 340 310 260 269 262

Impact on Social Perspective Taking

Play treatment 0.169** 0.239 0.227 0.026
(0.080) (0.143) (0.133) (0.114) Movie treatment -0.045 0.046 -0.241* (0.087) (0.104) (0.104)

Impact on Knowledge

Play treatment 0.101*** 0.062 0.121*** 0.067*** 0.122*** 0.237***
(0.016) (0.051) (0.024) (0.016) (0.020) (0.046) Movie treatment -0.015 0.045* -0.005
(0.027) (0.024) (0.048) N 1,442 341 310 260 269 262

Impact on Interest in Theater Consumption

Play treatment 0.127* 0.035 0.260* 0.329** 0.217 0.319
(0.069) (0.097) (0.146) (0.129) (0.150) (0.255) Movie treatment 0.084 0.250 0.106
(0.104) (0.155) (0.186) N 1,442 341 310 260 269 262

Impact on Interest in Theater Participation

Play treatment -0.109 -0.108 -0.051 -0.090 0.148 -0.373***
(0.082) (0.130) (0.182) (0.152) (0.238) (0.117) Movie treatment 0.008 0.236 -0.033
(0.120) (0.242) (0.121) N 1,442 341 310 260 269 262

Note. Standard errors—clustered within classes—are presented in parentheses. Coefficients for age, grade, gender, and class are not presented. The movie is *Twelfth Night* and *Peter and the Starcatcher*. * $p < .10$. ** $p < .05$. *** $p < .01$ (two-tailed).

magnitude and falls short of being statistically significant once we control for a pretreatment measure of SPT. Similarly, the effect, which was marginal in Table 2, also falls short of being statistically significant. Generally null effects of the movie treatment controlling for pretest measures of the outcomes with the possible exception of Content Knowledge. When we control for prior seeing a movie may increase understanding of the plot and vocabulary of the stories by .051 of a standard deviation, but that effect at $p < .10$. It is important to note that in controlling for pretreatment measures, we decrease our sample size. While the addition of controls improves the precision of our estimates, the loss of sample size decreases the precision even more while point estimates remain largely similar. For these reasons, we prefer the full sample models without controls for baseline measures.

Descriptively, it is interesting to note that female students tend to score higher on the Tolerance, SPT, Theater Consumption, and Social Perspective Taking measures, but these differences mostly disappear when controlling for pretreatment measures of those outcomes. Similarly, white

Tolerance and Content Knowledge outcomes, but that entirely disappears when controlling for pretreatment measures of those outcomes. The play or movie treatments had differential effects on students by gender or race/ethnicity.

Discussion

The experimental evidence presented here clearly shows that students can benefit from school field trips to see live theater. The effect is significant with respect to measures of Tolerance and Content Knowledge. Whether we control for pretreatment measures of outcomes or not, students show greater Tolerance as well as greater understanding of the plot and vocabulary of stories if they see live theater.

There may also be a benefit from seeing live theater for student ability to engage in SPT, but that effect falls short of statistical significance when controlling for a pretreatment measure of that outcome. We gain precision by controlling for pretreatment measures, but we also lose some precision because measures were only collected for the last two plays. On balance, we prefer the full model without controls for baseline measures, but we leave the results for readers to judge.

There is some indication that students randomly assigned to see live theater become more interested in frequenting the theater, but this effect is only marginally significant and disappears when controlling for the pretreatment measure of that outcome. Again, this loss of effect is likely due to the reduction in sample size and loss of precision. Student interest in participating in theater does not seem to be affected at all.

It is also important to emphasize that the movie treatment does not seem to have a robust effect on any of these outcomes. Going to see live theater produces benefits that cannot be produced by watching a movie instead. And the fact that students who received the movie treatment did not go on a field trip suggests that the effects we have observed are caused by the experience of watching live theater and not simply caused by leaving the classroom.

While this experiment demonstrates that live theater field trips cause an increase in Tolerance and perhaps in the related concept of SPT, we do not know why these effects were produced. Our best explanation is that theater is a window for students to a broader world. Exposure to that broader world helps their understanding and acceptance of that broader world, which is why we see increases in Tolerance and SPT. Plays may be more effective than movies in helping students understand and accept that broader world because we react differently to human beings acting out a story in front of us than we do to human beings on a screen. The in-person experience may create greater emotional connections. This pattern of results is consistent with research by Greene et al. (2014) that finds similar effects when students take field trips to visit an art museum.

It is educationally significant and a bit surprising that watching a movie is not a particularly effective way of conveying information. Watching a play is. Watching movies is an extremely common school practice, but it may produce little learning. Going to see a play is a less common but appears much more effective. This is especially surprising given that many films may be higher quality productions than plays. But as we saw in our experiment, even seeing a university play of *Twelfth Night* taught students significantly more about the plot and characters than watching a film with award-winning actors like Helena Bonham Carter and Ben Kingsley. The in-person experience, again, appears to trump the screen.

Of course, we were only able to observe effects 7 to 8 weeks after students saw the plays, so we do not know if these benefits are long-lasting. And while we saw consistent results across multiple plays, produced by different theaters, and involving different school groups, all occurred in one particular place of the country. So we cannot know with confidence that these benefits would be produced for other schools.

Nonetheless, this is the first major piece of experimental evidence on how viewing theater affects students and it shows significant benefits. If we eliminate or further cut these cultural experiences from schools, we should attempt to replicate this experiment in other locations and with different groups.

REFERENCES

- Baron-Cohen, S., Wheelwright, S., Hill, J., & Raste, Y. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, autism spectrum disorder or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(2), 241–251.
- Bowen, D. H., & Green, J. (2009). Learning to think critically: A visual art experiment. *Educational Researcher*, 43(1), 37–44.
- Ellerson, N., & McCord, R. (2009). *One year later: How it continues to impact school districts*. American Association of School Administrators. Available online at <http://www.aasa.org/uploadedFiles/Resources/files/OneYearLater%20FINAL.pdf>
- Gadsden, V. L. (2008). The arts and education: Knowledge generation, pedagogy, and practice. *Review of Research in Education*, 32(1), 29–61.
- Gehlbach, H. (2004). A new perspective on perspective taking: A multidimensional approach to cognitive development. *Educational Psychology Review*, 16(3), 207–235.
- Gehlbach, H., Brinkworth, M. E., & Wang, M. (2012). The social perspective taking process: What is another's perspective? *Teachers College Record*, 114, 1–29.
- Gehlbach, H., Brown, S. W., Loannou, A., Boyer, M. A., Hudson, N., Niv-Solomon, A., & Goldstein, T. R. (2008). Increasing interest in social studies: Social perspective taking and self-efficacy in stimulating simulations. *Contemporary Educational Psychology*, 33(1), 1–15.
- Goldstein, T. R., & Winner, E. (2012). Enhancing empathy and the-

ory of mind. *Journal of Cognition and Development*, 13(1), 19–37.

MAY 2018 253

Government Accountability Office. (2009). *Access to arts education: Inclusion of additional questions in education's planned research would help explain decreased for some students*. Washington, DC: Author. Available online at <http://www.gao.gov/assets/290/286601.pdf> Greene, J. P., Kisida, B., & Bowen, D. (2014). Cultural capital, teacher bias, and educational success: New twins. *Social Science Research*, 64, 130–144. Kisida, B., Greene, J. P., & Bowen, D. (2014). Creating cultural consumers: The dynamics of cultural capital. *Journal of Education*, 87(4), 281–295. Lacoë, J., Painter, G. D., & Williams, D. (2016). *Museums as classrooms: The academic and behavioral impacts of "school field trips"*. Paper, Sol Price School of Public Policy, University of Southern California. Lee, B. K., Patall, E. A., Cawthon, S. W., & Steingut, R. R. (2015). The effect of field trips on preK-16 outcomes: A meta-analysis of research from 1985-2012. *Review of Educational Research*, 85(1), 3–49. Ludwig, M., Boyle, A., & Lindsay, J. (2011). *Arts integration research through the lens of Every Student Succeeds Act*. Washington, DC: American Institutes for Research. Retrieved from <http://www.air.org/resource/review-evidence-arts-integration-research-through-lens-every-student-succeeds-act> Rabkin, N., & Hedberg, E.C. (2011). *What the declines mean for arts participation*. National Endowment for the Arts, Washington, D.C. Retrieved from: <https://www.arts.gov/sites/default/files/ArtsLearning.pdf> Ruppert, S. S. (2006). *Critical evidence: How the ARTS benefit student achievement*. National Assembly of State Arts Agencies. Retrieved from <http://www.nasaa-arts.org/wp-content/uploads/2017/05/critical-evidence.pdf>

254 EDUCATIONAL RESEARCHER

from <http://www.nasaa-arts.org/wp-content/uploads/2017/05/critical-evidence.pdf>

AuThORS

JAY P. GREENE, PhD, is a Distinguished Professor of Education Policy at the University of Arkansas, 200 Graduate Education Building, Fayetteville, AR 72701; *hh018@uark.edu*. His research focuses on culturally enriching field trips, school choice, and the effect of schools on noncognitive and civic values.

HEIDI H. ERICKSON is a Doctoral Fellow at the University of Arkansas, 200 Graduate Education Building, Fayetteville, AR 72701; *hh018@uark.edu*. She focuses on education policy and school finance with issues ranging from school choice to arts and values education.

ANGELA R. WATSON is a Distinguished Doctoral Fellow of Education Policy at the University of Arkansas, 200 Graduate Education Building, Fayetteville, AR 72701, arw007@uark.edu. Her research interests include arts-related field trips and informal learning, STEM gender gap, and school choice with a focus on

MOLLY I. BECK, MPA, is a Distinguished Doctoral Fellow at the University of Arkansas, 201 Graduate Education Building, Fayetteville, AR 72701. Her research interests include the educational outcomes of English Language Learners, academic and nonacademic outcomes from experiential learning, and retention.

Manuscript received August 25, 2017 Revisions received December 20, 2017 and January 17, 2018 /