Presenting School Choice Information to Parents: An Evidence-Based Guide

APPENDIX

ies INSTITUTE OF EDUCATION SCIENCES

NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE

NCEE 2019-4003

This page intentionally blank for double-sided copying.

Presenting School Choice Information to Parents: Technical Appendix

October 2018

Alyson Burnett
Ira Nichols-Barrer
Steven Glazerman
Mathematica Policy Research

Jon Valant
Brookings Institution

Meredith Bachman
Project Officer
Institute of Education Sciences

NCEE 2019-4003
U.S. DEPARTMENT OF EDUCATION

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

This page has been left blank for double-sided copying.

ACKNOWLEDGMENTS

This guide, and the study on which it is based, was a collaborative effort that benefited from the contributions of many people. The study team received useful input from its technical working group: Sujata Bhat, Rebecca Jacobsen, Cait Lamberton, Ellen Peters, Aesha Rasheed, Peter Bergman, Judith Hibbard, and Paul Teske. Eric Johnson provided valuable guidance on consumer behavior research and the design of the experiment. Andrew Gelman answered questions about Stan software.

The online information displays were designed and implemented by Tembo, Inc., led by Meg Towle and Michael Moore. GfK provided the survey sample used in the study.

At Mathematica, Phil Gleason provided expert review at every stage of the design and report. Lisbeth Goble navigated the data collection through approvals, and Jesse Chandler was instrumental in designing the survey and planning experimental logistics, and Mariel Finucane advised on statistical issues. The team received expert programming assistance in Stan from Daniel Kassler and quality assurance review by Shira Mitchell. Adrienne Jones provided additional research assistance and project management support. The guide was edited by Molly Cameron and Daryl Hall. Steven Malick and Kristin Hallgren read drafts of the guide. The documents were produced by Laura Sarnoski, William Garrett, and Jessica Coldren, with graphics support from Brigitte Tran.

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURES

This page has been left blank for double-sided copying.

"Presenting School Choice Information to Parents: An Evidence-Based Guide" was informed by a recent study examining how different ways of presenting school information to parents may affect how that information is understood, perceived, and used. This technical appendix provides a full description of the study's design, methods (including the analytic approach), and findings.

A. DESIGN OF THE EXPERIMENT

The study was an online randomized experiment designed to see how parents of school-age children respond to different ways of displaying school information. The goal was to determine what best helps them understand the information, find it satisfying and easy to use, and how different information displays might influence their selection of schools. The study addressed the following research questions:

1. Which format for displaying data on schools is best—numbers only, icons, or graphs?
2. Is it useful to show parents district averages on key measures of school performance for comparison purposes?
3. Is it useful to include parents' opinions of schools as an additional source of information for other parents?
4. What amount of information about each school should be shown in a summary of the choices available to parents?
5. Does it matter in what order the school profiles initially appear on the information display?
6. Can information be presented in ways that encourage parents to choose an academically higher-performing school?

1. Overview of the experiment and its outcomes

To answer these questions, the U.S. Department of Education's Institute for Education Sciences (IES) contracted with Mathematica Policy Research to design and carry out an experiment that tested how parents respond to being shown different variations of a school information display. The experiment had three parts, shown in Figure A.1: (1) an initial survey; (2) random assignment to one of the study's variations of the information display (these displays are referred to in this appendix as "treatments," and there were 72 treatments in the experiment, as described in Section A.2); and (3) an outcomes survey which measured the study's outcomes of interest.

The experiment's sample of 3,500 individuals was obtained from a market survey research firm that identified individuals who met the study's criteria (annual income below $40,000 and living with school-aged children) and were interested in participating. The sample was not statistically representative of any well-defined population since individuals volunteered to participate.

In part one, the individuals identified by the market research firm were directed to a website where they took an initial survey that included screener questions to confirm they met the study's inclusion criteria (as described in Section B.4) and background questions about their characteristics and past experience choosing schools. In part two, those who were eligible for the

study were randomly assigned to be shown one of the 72 variations of the information display.[1] The information display included a set of profiles for 16 elementary schools in a hypothetical school district constructed specifically for the study. Below (in Section A.2) is a description of what did and did not vary across the different displays. Examples of these displays can also be found throughout the guide.

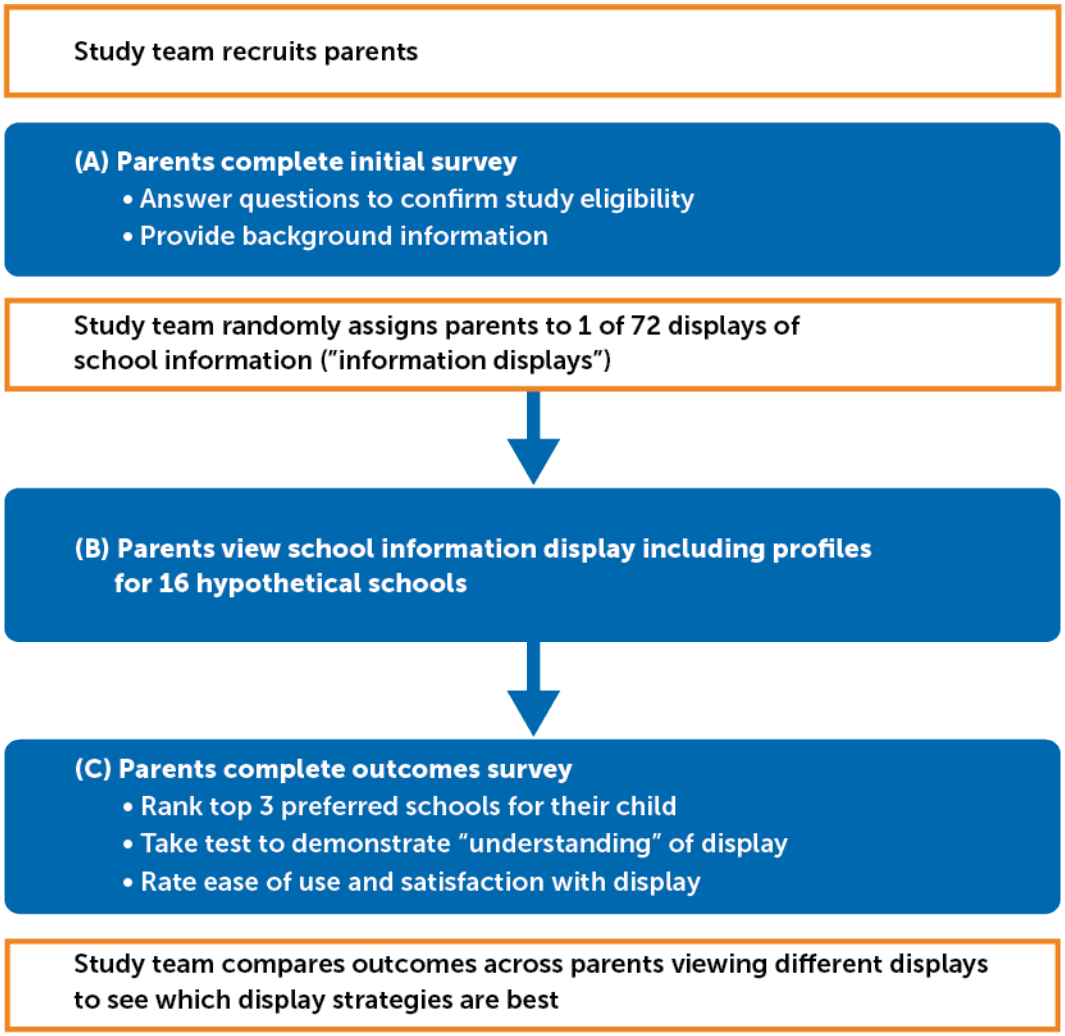
The third and final part of the experiment was an outcomes survey. This survey appeared in a separate web browser tab from the school information display, and participants were instructed to switch back and forth between the survey and the information display as they answered questions. The survey required participants to complete three types of tasks, each designed to measure one of the study's main outcomes.

- One task was for participants to select and rank their top three schools as if they had to choose one for their youngest child. This task was designed to measure effects of different types of information displays on choices.
- Another task was to answer factual questions about the schools in the hypothetical district, using the information display. This task was designed to measure impacts of different types of information displays on understanding.
- The third task was to agree or disagree with a set of statements about the experience of using the information display. This task was designed to measure usability; specifically, these statements were related to the ease of use of the information display and the participant's satisfaction with the display.

Section B.2 describes how the outcome constructs were defined.

[1] Eligible study participants were adults who had an annual income below $40,000, were living with children of school age, and who did not fail an "instructional manipulation check" in the initial survey that was designed to test if respondents were paying attention during the survey exercise. These criteria are discussed in detail in Section B.4.

Figure A.1. How the study worked



2. Design of the information display and its variations

The study was a randomized factorial experiment—one that examines several "factors." In this experiment the term factor refers to a specific aspect of how one presents school information to parents. For example, one factor varied was the **format** used to show parents data related to schools. It had three different display strategies: numbers, graphs, or icons. The strategies for each factor are illustrated in the guide and described and justified in more detail below.

To design the information displays used in the experiment, and select which specific display strategies to test and which to hold constant across all displays, the study incorporated insights from research in fields outside education and relied on the expertise of the design firm partnering with the study to create the information displays (this firm had experience creating school information displays for a variety of state and school district customers). Below, the design features that were included in all information displays are discussed first. This is followed by a discussion of the specific strategies that were tested.

- *Geographical map:* There was a map at the top of all displays, with a house symbol representing the parent's hypothetical home and markers representing the location of each school. Below the map was a list of 16 school profiles (16 rows of data, with one row per school) showing school information. Parents could also choose to conceal the map at the top of the display by pressing a "Hide Map" button. The prominent map was selected as a fixed feature in the study, because research has firmly established that proximity to home is an important consideration in determining parents' school choices (Glazerman 1998; Hastings, Kane, and Staiger 2009; Harris and Larsen 2015; Glazerman and Dotter 2017). Maps were also a standard feature on most school information displays examined at the time this study was conducted.

- *Standard set of information categories:* Each school profile in the display included information about the same four categories of information: distance from home, academics, safety, and resources. Research indicates that these four categories of information are often used by parents to make school selections (Glazerman 1998; Hastings, Kane, and Staiger 2009; Glazerman and Dotter 2017). However, this is not a comprehensive list of the categories of information that parents consider. Notably, the school profiles omitted data on student characteristics (e.g., race/ethnicity, free and reduced priced meals status). Although research indicates that parents often consider such information when choosing schools, doing so could have complicated interpretation of the study's findings.

- *Standard data layout:* Applying the advice and recommendations of the study's design experts, all information displays organized the school profiles using a standard organizational approach. Specifically, all of the displays structured the measures shown for each school in columns with fixed widths, with one column for each information category. Because each measure appeared under the same column for all schools, this approach makes it possible to scan vertically and compare distance from home, academics, safety, and resources for each school in the display. The study also did not test the effects of placing data on the left side of a website versus the right, as it is already well documented that readers in English-speaking countries tend to scan information from left to right and top to bottom (Hoekman 2010).

- *Glossary:* For each school information measure included in the display, there was a small question mark that parents could click to open a glossary explaining the definition of each school characteristic. The school information displays in this study contain terms such as "proficiency rate" that may be unfamiliar to many parents, or measures like "parent satisfaction" that require an explanation to understand how they were measured. Defining terms was important for this study, since participants came from every state in the country and might not have shared a common understanding of key terms due to differences in district and state policies, particularly those related to standardized tests.

- *Limited total number of measures shown for each school:* This was not highlighted as a common display feature in the guide (because the amount of information varied across displays), but the study did endeavor to limit the maximum number of measures that were shown to study participants. Specifically, the total number of measures shown for each school was limited to a maximum of 22 different measures in the higher-

information display. The study did not test the effect of including more than 22 measures, because the research literature in fields outside of education has established that there is a limit regarding how many pieces of information adults can process without becoming overwhelmed (e.g., Cronqvist and Thaler 2004). The maximum number of measures in this study was selected to make it possible for parents to see the map at the top of the display and all of the information about one school without needing to scroll up or down on a standard computer monitor.

These features were held constant to help the study focus on testing a specific set of display strategies. The study tested five factors—two with three strategies each, and three with two strategies each, meaning that the study tested a total of 72 different variations of information displays, or treatments (3 x 3 x 2 x 2 x 2 = 72). Each of these 72 combinations of strategies corresponds with a different variation of the school choice information display that differed in at least one way from all of the other displays in the experiment. The study randomly assigned each participant to one of these 72 variations of the information display. Because random assignment ensured there were no systematic differences between any of the treatment groups before the study began, the design allowed the study to treat the differences in outcomes between strategies as causal estimates of the effects of those strategies.

Table A.1. Factors and display strategies tested in the experiment

Factor	Strategy 1	Strategy 2	Strategy 3
1. Format	Numbers only	Numbers + icons	Numbers + graphs
2. District average	No district average	District average shown	n.a.
3. Source of information	District only	District + parent ratings	n.a.
4. Amount of information	Lower amount: one measure per category of information	Higher amount: multiple measures per domain all shown at once	Progressive disclosure: lower amount by default, with option to expand the view to the higher amount
5. Default sort order	By distance	By academics	n.a.

n.a. = not applicable.

Format (factor 1) was selected for the study because research outside of education suggests that the cognitive effort required to understand available choices can sometimes be reduced by applying visual design elements, such as graphs or icons, that may make it easier to understand information at a glance in a display (Nielsen 1989, 1993, 1999; Loranger and Nielsen 2006). In addition, past research outside of education indicates that the understandability and usability of an information display can be improved by more directly identifying when measures should be seen as positive or desirable, for example by using color-coding to show which values are better than others (Peters et al. 2009).

To examine how best to apply these principles in the realm of school choice, the study examined three different ways of formatting the display. One level uses a presentation style for academic and safety data that relies exclusively on numbers, a second level combines numbers with color-coded icons to label data in a readily identifiable way (letter-grade icons), and a third level combines numbers with graphical information (a horizontal bar chart showing a school's

academic and safety data relative to the maximum possible value) to facilitate visual comparisons of measures across schools. All displays included numbers. Typically, in school districts, information displays do include numbers alongside icons and graphs so pairing the graphical elements with numbers was more realistic and relevant to districts' practices. Also, the inclusion of numbers in each display was necessary to ensure that all participants could determine the correct answers to the study's understanding questions and to calculate the effects of different displays on parents' school selections in a standardized fashion (see discussion of outcomes below).

The study only tested one icon design. The study's single icon design (color-coded letter grades) was selected because strong evidence already exists in fields outside of education that color coding can directly influence choices (Singh 2006), and previous education research has also shown that letter grade labels can affect parental views of schools in some contexts (Jacobsen et al. 2014). For each measure where these letter grades were shown, grades were defined based on the relative ranks of the schools in the display: ranks 1-4 received an "A," ranks 5-8 received a "B," ranks 9-15 received a "C," and the school ranked 16th received an "F." Parents did not receive an explanation of how the grades were defined; however, the value of each measure (i.e., number) was shown next to each letter-grade icon, so it was possible to identify the cutoff value between grades by reading the information in the display.

District average (factor 2) was selected because past research outside of education has indicated that, in some circumstances, including one or more reference points in an information display helps to make the information more usable (Kahneman and Tversky 1979; Thaler 1985). For example, if a parent examining an information display was not sure how to interpret a measure of school performance, such as the proficiency rate on a standardized exam, it is possible that including a reference point showing if the school was above average or below average in the district might help to place the measure in context. On the other hand, including reference points could add unnecessary clutter to the display, or perhaps make the display less understandable to parents. To examine these issues, the study included a factor with two levels: one showing no district averages and another showing the district averages for academics, safety, and resources.

Source of information (factor 3) was included in the study because prior research has demonstrated that certain types of information gathered from parents can be more salient than district information in school choice decisions. That is, parents may prefer to use information from one another rather than rely on information provided by the district (Valant 2014). To examine this issue the study used a two-level factor, with one level exclusively presenting the district source of information (for example, test scores and suspension rates), and a second level combining the district source with parent satisfaction ratings. Valant (2014) has demonstrated that an especially salient type of parent information is "narrative comments"—written descriptions of a parent's experiences in a school, of the type that might be found in an email describing a school or an online discussion group for parents at a school. However, less evidence exists about the salience of parent survey ratings, which are more likely to be widely available on a district-wide basis (due to the financial and logistical challenges of collecting high-quality data from parents), and are therefore of greater relevance for the district guide.

Amount of information (factor 4) tested how higher and lower amounts of information affect the understandability and usability of the information shown to parents. Previous research outside of education has shown that limiting the amount of available information can both help and hinder choosers: it can improve the understandability of a presentation (Cronqvist and Thaler 2004) but can also remove the type of detail that could help choosers feel more confident in and satisfied with their choices (Bundorf and Szrek 2010; Chakraborty et al. 1994). To examine this trade-off, the study tested whether the amount of information shown at one time plays an important role in determining its understandability and usability. This factor varies the number of measures shown to study participants in three ways. The first level, "lower information," shows only a small number of measures in each school's profile—one measure for each category of information. The second level, "higher information," displays three to six measures for each information category. The third level, which is labeled "progressive disclosure" (but could be called "click to learn more") also provides access to this higher amount of information, but the additional data does not display by default. Instead, participants initially see the measures provided in the "lower information" strategy (that is, just one measure per category of information) and can click on an "expand" option in the school profile to reveal the extra details provided in the "higher information" display strategy.

Default sort order (factor 5)—varies how schools are sorted and numbered by default to examine the effects of organizing school profiles in different ways. Evidence that the sorting of data can influence choosers has been well documented in fields outside of education. Experiments have demonstrated that in some contexts, the initial sorting of options can lead choosers to place a disproportionate weight on the measure used to sort the data (Russo 1977; Cronqvist and Thaler 2004; Johnson et al. 2013). To date, however, no studies have investigated the effects of sort order on school choice websites and designs. In the experiment, sort refers to the order in which schools are displayed by default on the information display. For one display strategy in the study, schools are ranked by their distance from the participant's home, meaning that the closest school would appear at the top of the school list and the farthest school at the bottom. For the other strategy, schools are listed in order of academic performance, with the highest-performing school appearing at the top of the list regardless of its proximity to the participant. Each school received a number corresponding with its rank in this original sort scheme; these school numbers were used as a key corresponding to school markers on a district map shown prominently on the information display. Although participants could re-sort the data to list the schools in a different order, including alphabetical order, these original ranking numbers remained as a persistent reminder of the original default rankings shown to them.

3. School information to display

The study asked parents to consider a hypothetical district with 16 elementary schools, with school information designed by the research team. The research team generated information for each school regarding: (1) location, (2) academic performance, (3) school safety, and (4) resources. These categories of information were selected after reviewing a wide range of districts' school information displays and identifying the types of information that districts most commonly make available to parents. Parents also report that these categories of information are among the items they care about most when selecting schools (Valant 2014).

Table A.2 summarizes the measures shown under each category of school information and illustrates how they varied based on each of the factors discussed above. As shown in the table, the study did not vary the format or source of data for measures related to distance or resources, and the district average was also not shown for distance measures. The table also lists which measures appeared only in the higher-information versions of the display, and which measures were used as parent ratings of school academics and safety.

Table A.2. Variations in the school measures appearing in tested displays

Category of information	Amount of information*	Format (icons or graphs shown in addition to numbers)	Source (parent ratings shown in addition to district measures)	Default sort order	District average
Distance	Lower amount: straight-line distance from home to school Higher amount: walking time, driving time	No variation: distance always shown using numbers	No variation: only district measures were shown	By distance: schools sorted by distance from home to school By academics: schools sorted by % proficient on 2016 achievement test	No variation (district average never shown)
Academics	Lower amount: % proficient in both math and reading on 2016 achievement test Higher amount: % proficient on the 2016 math test, % proficient on the 2016 reading test, average 2015–2016 academic growth (0–100 index), average 2015–2016 academic growth in math, average 2015–2016 academic growth in reading	Numbers only*:* one number per measure Icon*:* Letter-grade icon next to number with color coding (green indicating better grades) Graph*:* Horizontal bar graphic next to number for each measure	Parent rating: percentage of parents agreeing with statement that they are highly satisfied with the school's academic quality	By distance: schools sorted by distance from home to school By academics: schools sorted by % proficient on 2016 achievement test	District average: District average shown for all included measures

Category of information	Amount of information*	Format (icons or graphs shown in addition to numbers)	Source (parent ratings shown in addition to district measures)	Default sort order	District average
Safety	Lower amount: % of students with no suspensions Higher amount: attendance rate, school won a blue-ribbon award for anti-bullying efforts (yes/no)	Numbers only: one number per measure Icon: Letter-grade icon with color coding for suspensions and attendance measures Graph: Horizontal bar graphic for suspensions and attendance measures	Parent rating: percentage of parents agreeing with statement that the school is a safe place for their child	By distance: schools sorted by distance from home to school By academics: schools sorted by % proficient on 2016 achievement test	District average: District average shown for all included measures
Resources	Lower amount: number of laptops or tablets per 100 students Higher amount: year of most recent school renovation, school has dedicated art studio, library, computer lab, or music program (yes/no)	No variation: resources always shown using numbers	No variation: only district measures were shown	By distance: schools sorted by distance from home to school By academics: schools sorted by % proficient on 2016 achievement test	District average: District average shown for all included measures

* Measures included in the "lower amount" displays were also present in the "higher amount" displays.

After examining a variety of existing school information displays and consulting with experts in school selection processes and decision making, the research team decided to include a total of 16 schools in the information displays used in the study. The number of included schools was selected to support the study's research goals while still appearing realistic when considering the size of many districts with information available online. The values of the measures shown for each hypothetical school were generated using several principles:

- No one school was uniformly the best for all four categories of information. The study ensured this would be the case, to force parents to make trade-offs when selecting schools. This allowed the study to measure the degree to which different factors influenced the types of schools parents might choose.
- One school was uniformly the worst. The study created a dominated school that was an extreme outlier in all information categories. It was farthest from the home symbol, had very low test score proficiency, and so on. This school was listed first in alphabetical order to help identify inattentive study participants. Anyone who selected the dominated school as one of their top three was assumed to be an insincere participant and was removed from the analytic sample. The frequency of these cases is discussed in Section B.4.

- The study ensured that within each of the four categories of information, the addition of higher amounts of information did not change the rank ordering of the schools. This was done because the study sought to measure the pure effect of adding more information without introducing ambiguity regarding which of the measures influenced the parents' school preferences.
- To generate values for the measures shown for each school, the research team began by assigning a master rank to each school in each of the four information domains, and then used those ranks to derive a value for each measure in each school (based on a pre-determined range and standard deviation selected for each measure). The research team then added random jitter to these values, to make the detailed information look realistic. This adjustment to the values of the measures—for example, adding or subtracting small numbers from the proficiency rates for math and reading—was small enough to maintain the rank ordering of schools but also make it difficult for study participants to figure out that the more detailed information did not change any of the relative rankings.

B. METHODS

1. Bayesian analysis

The study used a hierarchical Bayesian regression model. This kind of model looks much like a classical linear regression model, in that it estimates an effect size for each display strategy, estimates interaction effect sizes that make it possible to predict the effectiveness of each pairwise combination of strategies, and includes controls for participants' demographic characteristics to improve statistical precision. However, Bayesian analysis differs from the classical frequentist analysis used with linear regression models in important ways. First, instead of estimating whether a single null hypothesis is true (i.e., whether the analysis should reject the hypothesis that there were no significant differences between strategies), the Bayesian approach allows the researcher to calculate the **posterior probability** that the average outcome is better for one strategy than another or set of others. Second, Bayesian analysis provides more statistical precision (through a process known as "partial pooling" or "shrinkage adjustments") since treatment arms in a hierarchical Bayesian analysis are able to borrow strength from observations in other related treatment arms. This is especially valuable in an experiment such as this one, with a relatively small sample of participants assigned to each treatment arm (with 72 treatment arms and 3,500 participants, there were approximately 49 participants per arm). Another advantage of partial pooling is that the study does not have to ignore or risk overadjusting for multiple comparisons (see Gelman 2012).

a. Statistical model

In more formal terms, the impact model estimated 12 main effects (one effect for each level of the study's five factors), and pairwise interaction effects between strategies. To make the factorial design more efficient and to simplify the results, the design focused only on two-way interactions. That is, third-order and higher interactions were assumed to be zero. Each participant, *i*, was randomly assigned to a treatment arm defined by the following five experimental factors:

1. *g*, the format: use of numbers, graphs, or icons in the information display (a three-level factor)
2. *n*, the presence of district averages as a reference point (a two-level factor)
3. *s*, the source of the information (a two-level factor)
4. *l*, the amount of information (a three-level factor)
5. *o*, the default sort order of the schools (a two-level factor)

The study analyzed data from participants in all 72 treatment arms to estimate the following model:

$$y_i = \alpha + \sum_{m \in F} \beta_{m_i}^{(m)} + \sum_{q,r \in F, q \neq r} \theta_{q_i,r_i}^{(q,r)} + \gamma X_i + \varepsilon_i$$

In the equation above, participants are indexed by *i*, so that y_i is the outcome of interest for participant *i*. Before the model was estimated, each outcome was standardized to have a mean of

0 and a standard deviation of 1. The decision to standardize each outcome in this way is a recommended practice in Bayesian analyses because it makes it easier to compare the priors selected for the analysis across multiple outcomes with different scales, and benchmark those decisions against a wide range of other studies that standardize outcomes using a similar convention (Gelman and Hill 2007). The set $F=\{g,n,s,l,o\}$ represents the set of factors given above, and it is used in two ways: the index (m) indicates a choice of one of the 5 factors $m \in F$, whereas the index m_i indicates the level of factor m for participant i. To make this explicit, the effects of factor m are shown in the model equation as a vector $\beta^{(m)} = \left(\beta_1^{(m)}, \ldots, \beta_{J^{(m)}}^{(m)}\right)$ where $J^{(m)} = \{2,3\}$ is the number of levels of factor m; the term $\beta_{m_i}^{(m)}$ is the main effect of factor m at level m_i. For example, if a participant i was randomly assigned into a treatment arm with district averages not shown in the display ($n_i = 1$), the main effect of this display feature would be $\beta_1^{(n)}$ for that participant. Interaction effects are represented similarly: the term $\theta_{q_i,r_i}^{(q,r)}$ represents the interaction effect of factors $q, r \in F$ for each possible pair of strategies, q_i and r_i, that a parent could receive. For example, estimated interaction effects will allow the effect of the graphs strategy to vary depending on whether the amount of displayed information is low (one measure per information category), higher (multiple measures for each category), or higher with progressive disclosure. The vector X_i is a set of binary covariates controlling for participant characteristics (household income, parent age, parent education level, parent internet use, parent experience with school choice, and child gender, child race, and child special education status), with effects given by γ. Finally, α is an intercept, and ε_i is a participant level error term.

The model described above is a hierarchical Bayesian regression model, meaning that the analysis assigns a hierarchical set of "prior" distributions to the relevant parameters in advance, before estimating effects. These prior distributions were as follows:

$$
\begin{aligned}
\beta^{(m)} &\sim \mathcal{N}\left(0, \tau^{(m)}\right) \\
\theta^{(q,r)} &\sim \mathcal{N}\left(0, \tau^{(q,r)}\right) \\
\epsilon &\sim \mathcal{N}\left(0, \sigma\right) \\
\tau^{(m)} &\sim \mathcal{N}\left(0, \phi_{\text{main}}\right) \\
\tau^{(q,r)} &\sim \mathcal{N}\left(0, \phi_{\text{int}}\right) \\
\alpha, \sigma, \gamma, \phi_{\text{main}}, \phi_{\text{int}} &\sim \mathcal{N}\left(0, 3\right)
\end{aligned}
$$

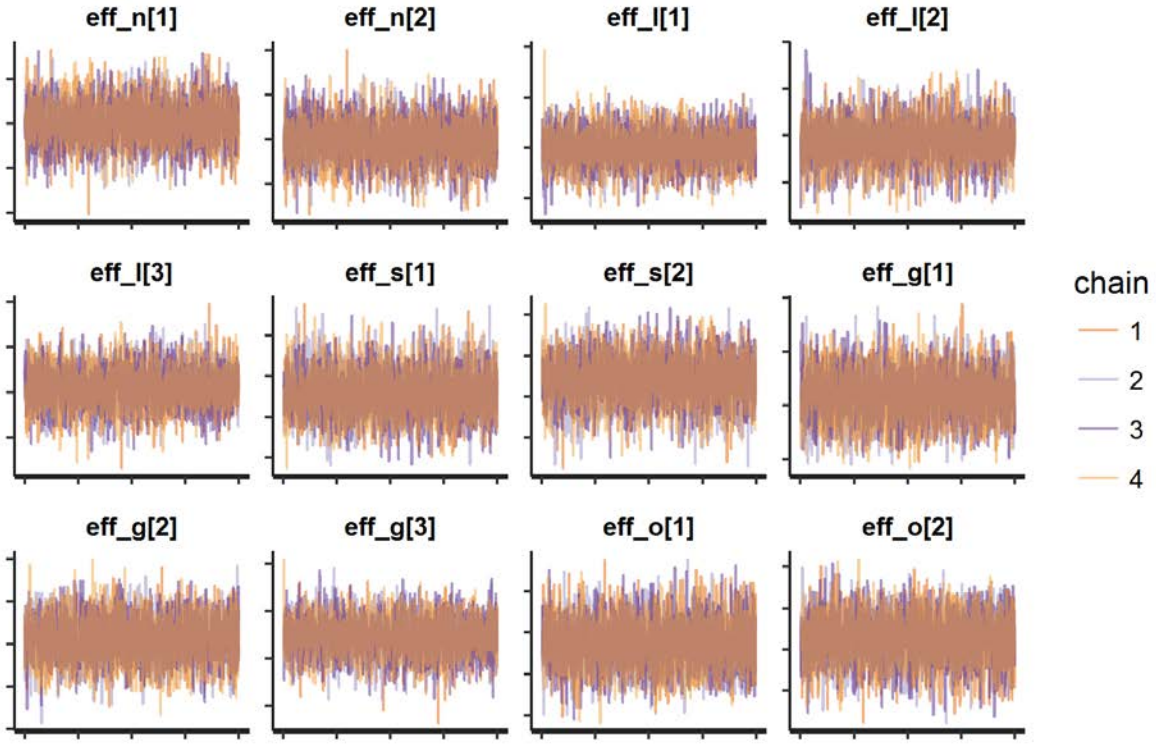
This list shows that each main effect and interaction effect has a separate prior. These priors are defined in advance by a mean (the first value in each set of parentheses, which is always zero in this model) and a standard deviation (the second value in each set of parentheses, which varies for each parameter). Using priors with a mean of zero indicates that the study was neutral prior to the experiment regarding which main effects and interaction effects will be positive or negative. The model is hierarchical in the sense that the standard deviation of main effects and the standard deviation of interaction effects (none of which are specified in advance of the experiment) are in

turn defined by their own, separate normal priors (the priors for these variance terms are "half-normal," meaning that they cannot take negative values). This use of hierarchical "hyperpriors" provides the model's framework for partial pooling across treatment arms (Gelman 2006).

The choice of a normal probability model to estimate average treatment effects (that is, using a model where the errors are assumed to be normally distributed) is a common practice in both Bayesian and frequentist regression analyses. This type of model is commonly recommended whenever a study is seeking to estimate the average treatment effect for a given sample, both for binary outcomes and non-binary outcomes with a wide range of potential distributions. Discussions of the merits of these models when estimating treatment effects can be found in Deke (2014) and Gelman and Hill. (2007).

To estimate the effects of each display strategy, the study used the Stan programming package. The Stan package used a Hamiltonian Monte Carlo (HMC) algorithm (a variant of Markov Chain Monte Carlo algorithm) to estimate the posterior distribution the hierarchical model in this study. To ensure the validity of the study's fitted models, for each of the analyses presented in the guide the study carefully examined the performance of the HMC process used by Stan to estimate effects for each model, and these diagnostics included several convergence checks. The thresholds used by the study for accepting estimated results were consistent with the default standards established in the Stan programming language and recommended by the Stan development team. Specifically, in estimating each of the study's core outcomes (effects for the full sample on understanding, ease of use, satisfaction, and the characteristics of schools selected by parents) no parameters of the model had a Gelman-Rubin potential scale reduction factor (R-hat) above 1.1, no iterations of the sampler encountered divergent transitions or exceeded the maximum treedepth, and in the sampler results no parameters had an effective sample size less than 10 percent of the total sample size and no parameters had a Monte Carlo standard error greater than 10 percent of their posterior standard deviation. Finally, the study performed visual inspection on traceplots of the key model parameters to ensure the separate chains of the sampler were well mixed for each of the estimated effects in the study (a sign that the sampler has appropriately explored most of the posterior distribution). An example of these traceplots is shown in Figure B.1., for the model estimating the impacts of each display strategy on the understanding outcome—traceplots for the other results presented in the guide were similarly well-mixed.

Figure B.1. Model estimation traceplots for the understanding outcome



Note: These figures show traceplots for each of the 12 main effects of the model (the estimated average effect of each display strategy, using the regression model's notation to label each factor and factor-level). Each plot indicates the values taken by the understanding outcome (vertical axis) over the 1000 iterations of the Stan software package's Hamiltonian Monte Carlo sampler (horizontal axis) for the four independent chains of the sampler (colored lines). If the chains of the model are well-mixed the paths shown in the four traceplots should frequently change direction and rarely stay in the same place, and the four chains should all occupy the same range in the parameter space. These traceplots are all consistent with a model that has converged to and efficiently explored the true posterior for all 12 main effects. These figures show traceplots for the understanding outcome: traceplots for the models estimating effects on the study's other outcomes (ease of use, satisfaction, and effects on choices) were similarly well-mixed.

b. Posterior probabilities

The study's results can be used to calculate the probability that a given strategy is better than the other strategies tested for that factor in the experiment (with respect to the outcomes measured). The guide describes a given strategy as having outperformed its alternatives (the other levels of the same factor) when the posterior probability of producing the highest score on the outcome in question is 70 percent or higher. For example, for the format factor, the guide reported that the display using graphs was best for satisfaction because the analysis suggested that there is a 71 percent probability that the graphs display performed better (resulted in a higher average value) on the satisfaction outcome than the numbers-only display and the numbers-plus-icons display. The study also examined whether each strategy had an effect on the characteristics of schools parents selected on behalf of their child. To measure if a positive effect occurred, the analyses estimated whether 70 percent or more of the posterior distribution for that effect fell above a value of zero.

The 70 percent posterior probability threshold was informed by guidelines established in the literature on Bayesian decision analysis (e.g., Berger 1985; Gelman et al. 2013; Isakov et al. 2015; Goodman 1999a; Goodman 1999b). Bayesian decision theorists argue that some types of decisions are justifiable only when there is a high degree of confidence, but other types of decisions are justifiable with lower degrees of confidence. When the consequences of making a mistake are very large—for example, in a medical trial that tests the effectiveness of an intervention that is expensive and that has serious side effects—it would be appropriate to require a high degree of confidence that the intervention works (95 percent or higher) before making a recommendation.

This study, in contrast, is testing alternative ways of displaying information that in general have low incremental costs. Indeed, the costs of implementing the tested levels of each factor are, in many cases, effectively equivalent. For example, changing the sort order of schools or including district averages in an information display could be done at little cost compared with the alternatives that were tested (although collecting parent survey data or redesigning the presentation format of a display may be more costly than the other factors tested in the study). In contexts such as these, Bayesian decision analysts agree that confidence thresholds of 70 percent or even lower than 70 percent are appropriate (Berger 1985). For the sake of simplicity, the study used the same 70 percent probability cutoff for findings regarding both two-level and three-level factors.

c. Subsample analyses

The study also examined whether the impact findings for the overall sample of low-income parents were consistent with findings for subsamples of parents with particular characteristics: parents with school choice experience, parents with lower education levels, parents with lower incomes, and parents who do not use the Internet intensively. The study defined parents with school choice exposure as those who reported having the option to choose a school without moving or who applied to a non-neighborhood-assigned public school (n = 1,819). Parents with lower education levels were those without a postsecondary degree (n = 2,068), and lower-income parents were those with annual household incomes below $30,000 (n = 2,241). Non-intensive Internet use was defined as less than 30 hours per week (n = 2,498)—this cutoff was selected to remove the heaviest Internet users from the sample. The subsample results are discussed in Section C of this appendix.

It was necessary to define rules by which subsample findings should be considered consistent or inconsistent with full sample findings. These rules recognize the fact that subsamples reduce the size of the sample, and therefore reduce the statistical precision of the estimates. Reduced statistical precision means that even when the true effect for a subsample is equivalent to the overall sample, a finding could "change" in the sense that the posterior probability of a strategy being best falls below the 70 percent threshold discussed above. To account for this, the study adopted consistency rules that recognize when there is evidence that a policy recommendation made for the full sample should change for the subsample. Specifically, the study used two consistency rules, illustrated in Table B.1.

1. **Overturned non-recommendation.** If the full sample analysis did not result in a recommended strategy for a given factor and outcome but the subsample did (i.e., the

probability that any strategy was best exceeded 70 percent in the subsample but not the full sample), then the findings were considered inconsistent.

2. **Different best strategy.** If the full sample analysis resulted in a recommendation for a given factor and outcome, then the subsample results were considered inconsistent if a different strategy had the largest effect size for that factor and outcome (regardless of whether the posterior probability in the subsample exceeded the 70 percent threshold).

Table B.1. Criteria for determining if subsample results were consistent with the full sample, for a given factor and outcome

		Subsample: Was any strategy more than 70% likely to be the best?	
		No	Yes
Full sample: Was any strategy more than 70% likely to be the best?	No	*C*	*X*
	Yes	Same strategy had the largest effect size: *C* Different strategy had the largest effect size: *X*	

Note: *C* Results are considered consistent. *X* Results are considered inconsistent

d. Sensitivity tests

In addition to subsample analyses, the research team also performed a set of sensitivity tests to examine whether the main pattern of results holds up if different methodological decisions had been made (e.g., constructing outcomes, sample screening criteria, and statistical modeling). This provides a sense of how "robust" the guide's findings are. For each sensitivity test, the study examined whether the pattern of conclusions drawn from the sensitivity analyses differed from those of the study's primary model. That is, for each outcome and sensitivity test, the study examined whether the alternative model identified the same strategies as being best for each factor, compared with the primary model. The results are described in Section C.2.

2. Approach to measuring outcomes

This section and the next one describe how the research team constructed each outcome measures and the rationale behind specific decisions. The study also examined alternative outcome measures as sensitivity tests, which are described in Section C.2 of this appendix.

a. Background on the survey items used for outcome measurement

The outcomes in this study include measures of understanding, usability, and effects on choices. The literature on school choice suggests that information about schools is often too complex for parents to understand and use for decision making (Schneider and Buckley 2002). Thus, a goal for many districts producing school information displays—made explicit in ESSA's report card requirements and regulations—is to make sure that the information is presented in a user-friendly and understandable manner. Furthermore, research on judgment and decision making shows that the way information is presented can influence the characteristics to which people pay attention and encourage them to make certain choices (for example, Castleman and

Page 2015; Thaler and Sunstein 2009). For this reason, it was important to test how the information displays would influence the characteristics of the schools that parents may select.

In developing the survey items used to assess the outcome measures, the study sought to draw from techniques commonly used in related fields and existing validated measures.

- **Understanding.** To measure understanding of the information shown in the 16 school profiles, the study included a comprehension test with items assessing whether participants could correctly identify facts about the schools. The use of a comprehension test was motivated by research showing that correlations between self-reported and actual knowledge and abilities are often quite poor (e.g., Kruger and Dunning 1999). The study used a pilot survey to test a list of potential items and, for the final measure, eliminated redundant ones, in the sense that they did not further discriminate between high and low understanding given the other items in the module. The study also eliminated items that only a small number of participants answered correctly.
- **Usability.** Usability was defined as the extent to which participants perceive that information is satisfying and easy to use. To assess usability, the study drew from the System Usability Scale, or SUS (Brooke 1986), a reliable and validated measure that has been used in more than 1,300 articles and publications. The SUS was supplemented with items tailored to assessing the usability of an information display for selecting schools. Specifically, the items added to the SUS asked parents to rate their confidence with their understanding of the school profiles and willingness to recommend the displays to a friend. After the pilot survey data was collected, factor analyses revealed that the combination of SUS survey questions and newly developed study questions formed two different outcome indices: one 'ease of use' index measuring how much effort was needed to use the school information display; and one 'satisfaction' index measuring how parents felt about using the information display to select schools for their youngest child. These two factors are in line with literature outside of education that suggests satisfaction and ease of use are distinct concepts (Brooke 1986)—as a participant may find a display easy to use because it is simple but be dissatisfied because it does not show enough information. The exact items in each index can be found below, in Table B.3.
- **Effects on choices.** To measure the effect of school information displays on the characteristics of selected schools, the study included a type of selection task that is commonly used in other studies of choice decisions (Akaah and Korngaonkar 1983; Jansen et al. 2000; Lewinsohn and Mano 1993; Srinivasan and Park 1997). After seeing the information display, parents completed a two-step process of first selecting their top 3 schools (the preferred schools they might select on behalf of their youngest child) from the list of 16 and then ranking the top 3 from most to least preferred. The schools that parents selected revealed the importance of various school characteristics—such as distance from home, academics, safety, and resources—in their decision making.

Before launching the survey, the research team vetted all items with an expert consultant and conducted two separate pilots of the survey.

b. Selection of outcome scales

In constructing outcome indices, a top priority was to make sure that readers of the district guide could easily understand and interpret the outcomes. Once the study identified a measure that was grounded in theory and could easily be interpreted, the research team compared it to alternative scale constructions by examining the following properties:

- Descriptive statistics (for example, minimum, maximum, mean, median, standard deviation, and variance)
- Floor and ceiling effects (through examination of histograms)
- Internal consistency reliability (Cronbach's alpha)
- Comparison of fit of competing factor analytic models

3. Construction of outcome measures

a. Understanding

Survey items. The module that tested parents' understanding of the factual school information consisted of six question items. The first four items required the participant to choose one out of two named schools that met a single criterion. The last two items required the participant to choose two or more schools from a list of five schools that met multiple criteria. The items were as follows:

- Which school is closer to the home icon on the map? (1)
- Which school has better academic performance as measured by standardized test scores? (2)
- Which school has better student discipline? (3)
- Which school has more access to laptops and tablets? (4)
- Select the three schools from the list below that are (5):
  - No more than one mile from home
  - Also have at least 50 laptops or tablets per 100 students
- Select the two schools from the list below that are (6):
  - No more than 1.5 miles from home
  - Also have an academic proficiency rate of at least 75
  - Also have at least 90 percent of students that have never been suspended

Index construction. The study constructed the understanding index based on the accuracy of responses to each item in the module. The number of points awarded for a correct response reflected the number of school criteria that needed to be evaluated. Thus, the study awarded one point for each correct response to items 1–4, two points for item 5 (no partial credit), and three points for item 6 (no partial credit). Giving greater weight to more complex items is a standard practice in educational assessment. Partial credit was not awarded for items 5 and 6 because with

five response options, there was a high probability (20 percent) that a participant would select a correct school by chance. Raw scores on the understanding index range from 0–9. Table B.2 provides descriptive statistics for the resulting outcome index: the distribution was approximately normal, but the values are right-censored because the modal score on the understanding index was close to the maximum value of 9. To make the scores more interpretable, the study converted this 0–9 scale to a 0–100 scale for the guide, by dividing the raw score by 9 and multiplying the result by 100. The resulting score represents the percentage of points awarded for correct answers.

Table B.2. Descriptive statistics for the understanding outcome

Index	Minimum	Maximum	Mean	Standard deviation	Cronbach's alpha
Understanding	0.0	9.0	6.6	2.7	0.65

b. Usability

Survey items. The survey included two modules on usability, each of which consisted of a series of statements for which the study gave participants a four-point agreement-disagreement scale. The first module was administered after participants completed the school selection task (that is, selected their top three schools). The second usability module occurred after participants completed the understanding module (that is, the test). The statement wording for each of these modules and the sources for each item are shown in Table B.3. Most of the items were adapted from the SUS. All other items were developed to address aspects of usability specific to the school information displays. Item correlations are shown in Table B.4 and descriptive statistics are shown in Table B.5.

Responses for all usability items ranged from 1 (strongly disagree) to 4 (strongly agree), with each possible response value labeled. The study scored each item as 1 (agree or strongly agree) or 0 (disagree or strongly disagree). Some items had a negative valence (2_2, 2_4, and 2_8), meaning that the statements were negative toward the information displays ("too complex," "I needed help," "I was unfamiliar with the terms"). These items were reverse coded so that a value of 1 could always be interpreted as more usable. The total score is the average across the items included in the index, such that scores range from 0 to 1. This method results in units that are easy to interpret as the proportion of items with which participants agreed.

Table B.3. Items included in satisfaction and ease-of-use scales

Module	Item wording	Code	Source	Satisfaction	Ease of use
1_1	It was easy to choose my top 3 schools.	easy-choose	New		
1_2	It was easy to decide how to rank my top 3 choices.	easy-rank	New		
1_3	I am confident that I understand the information in the school profiles well enough to pick the best school.	confident-pick	SUS	*	
1_4	The school profiles gave me enough information to pick the best school.	enough-info	New	*	
1_5	I would recommend browsing these school profiles to a friend who was also trying to shop for schools.	recommend	New	*	
2_1	If I had to choose a school for my child, I would like to use a set of profiles like this one.	like-to-use	SUS	*	
2_2	The school profile information was too complex. (*reverse coded*)	info-not-complex	SUS		*
2_3	The school profile information was easy to use.	info-easy-to-use	SUS		*
2_4	I think that I would need someone to help me use this set of profiles effectively. (*reverse coded*)	no help-needed	SUS		*
2_5	Most people would learn to use these school profiles very quickly.	Learn-quick	SUS		*
2_8	I was unfamiliar with many of the terms used in the school profiles. (*reverse coded*)	no-unfamiliar-terms	New		*

Note: Items 1_1 and 1_2 were not included in either scale because responses may have been influenced by how parents felt about the schools described in the information displays versus how they felt about the displays themselves. Items 2_6 and 2_7 were deleted after pilot testing and not included in the final survey. The SUS is the System Usability Scale. Where the item source is "new," it was developed to measure usability issues that are specific to a school information display.

Table B.4. Correlations, means, and standard deviations of usability items on a 0–1 scale

	1_1	1_2	1_3	1_4	1_5	2_1	2_2	2_3	2_4	2_5	2_8
1_1	1.000										
1_2	0.525	1.000									
1_3	0.329	0.351	1.000								
1_4	0.295	0.231	0.447	1.000							
1_5	0.316	0.330	0.473	0.433	1.000						
2_1	0.217	0.207	0.227	0.237	0.358	1.000					
2_2*	0.021	0.045	0.046	-0.029	0.039	0.152	1.000				
2_3	0.176	0.206	0.185	0.142	0.206	0.455	0.315	1.000			
2_4*	-0.006	0.038	0.051	-0.013	0.042	0.085	0.584	0.181	1.000		
2_5	0.154	0.169	0.159	0.169	0.225	0.378	0.170	0.455	0.146	1.000	
2_8*	0.009	0.028	0.067	-0.006	0.002	0.066	0.539	0.134	0.547	0.091	1.000
Mean	0.845	0.897	0.917	0.823	0.927	0.936	0.829	0.933	0.781	0.929	0.811
SD	0.362	0.305	0.276	0.382	0.260	0.244	0.376	0.250	0.414	0.257	0.391

Note: Darker shades (larger numbers) illustrate higher correlations

*Items 2_2, 2_4, and 2_8 were reverse coded so that all items would have a positive valence.

Table B.5. Descriptive statistics for usability measures

Index	Number of items	Minimum	Maximum	Mean	SD	Cronbach's alpha
Satisfaction	4	0	1	0.90	0.21	0.68
Ease of use	5	0	1	0.86	0.23	0.71

Index construction. The study team considered three factor structures and compared the fit of each using confirmatory factor analysis (CFA). Each model had two factors. The results are reported in Table B.6.

Table B.6. Fit of alternative factor structures

	Model 1*	Model 2	Model 3
Factors	Satisfaction: 1_3-5, 2_1 Ease of use: 2_2-5, 2_8	Satisfaction: 1_3-5, 2_1 Ease of use: 1_1-2, 2_2-5, 2_8	Module 1: all items Module 2: all items
SRMR	0.15	0.14	0.11
RMSEA	0.16	0.19	0.16
Bentler CFI	0.77	0.66	0.77
Cronbach's alpha	Satisfaction: 0.68 Ease of use: 0.71	Satisfaction: 0.68 Ease of use: 0.66	Module 1: 0.74 Module 2: 0.71

Note: SRMR = Standardized Root Mean Square Residual. Lower values indicate better fit.
RMSEA = Root Mean Square Error of Approximation. Lower values indicate better fit.
CFI = Comparative Fit Index. Higher values indicate better fit.

*Selected model, described in Table B.3.

The study ultimately selected Model 1 and formed two separate usability indices—satisfaction and ease of use, as shown in Table B.3. Although Model 3, which combined items

related to satisfaction and ease of use, had slightly better fit in some respects, the results of the factor analysis were just one consideration for determining how to form the usability indices. For the purpose of the guide, it was more important to ensure that the indices were interpretable and represented clearly distinct outcomes for users of the guide.

c. Effects on choices

The study measured the effect that information displays might have on eventual school selections by having participants review the set of 16 school profiles and select their top 3, ranking them in preference order as if they were choosing a school for their own youngest child. This exercise is hypothetical and thus not the same as the high-stakes choices made for children applying to real schools, but it does provide initial evidence that could generate hypotheses for a field experiment. In the study, choices revealed the weight parents placed on various categories of school information to make their selection. For example, if parents chose a school that had the highest possible standardized test scores but a small number of laptops per 100 students, it would suggest that in their school choice they were placing more weight on the school's academic achievement than its resources.

To score the effects on choices, the study measured the average characteristics of the three schools selected by parents, weighting the higher-ranked of the three selected schools more heavily. Specifically, the top-ranked school received double the weight of the second-ranked school, and the second-ranked school received double the weight of the third-ranked school. For each measure, the schools' characteristics were converted to a standardized *z*-score value (mean of zero and standard deviation of one across all schools in the data) for the four measures present in all of the displays: distance (miles from home), academic quality (percentage proficient on the state exam), safety (percentage of students with no suspensions), and resources (laptops per 100 students).

4. Sample screening, survey protocol, and randomization

This section summarizes key information about the study sample: the selection criteria, demographic characteristics, and results from randomization, including attrition and baseline equivalence.

a. Screening criteria and survey integrity

The sample was an opt-in, nonprobability sample in which a provider of online samples recruited potential participants and ran the survey until a quota was met. Table B.7 shows how the study sample was formed. A total of 6,160 potential participants were screened into the initial survey by the sample provider and began the survey, 3,961 passed the screening checks in the initial survey and reached the point of random assignment, and of those who were randomly assigned the study included 3,500 in the final analysis due to attrition, after applying additional criteria to screen for participant effort (discussed below). The steps of the survey are illustrated in Figure A.1, at the beginning of this appendix.

All of the potential participants were adults who had previously signed up to complete surveys with the provider in exchange for compensation and had answered detailed questions about their employment, households, and children. The provider then used its sample database to identify those who met the study's selection criteria: (1) an annual household income of $40,000

or less; and (2) at least one child in grades kindergarten through 12. The initial survey also included questions that verified potential participants' income, the grades of their children, and included an additional screener question asking if the participant was involved in educational decisions for their youngest child. Prior to random assignment, the study screened out those who did not meet the selection criteria and could not continue the survey (10 percent of those who agreed to participate).

The survey also included an "instructional manipulation check" (Oppenheimer et al. 2009) prior to random assignment to encourage potential participants to read instructions carefully and filter out those who were disinclined to read instructions. The survey included an item that had both a seemingly obvious but incorrect answer and a correct answer that could be determined only after reading the instructions. Potential participants were allowed four attempts to complete this item. Fourteen percent of those who attempted the instructional manipulation check (IMC) did not continue past the item; specifically, 11 percent were screened out because they responded incorrectly after four attempts, and 4 percent quit during the task.

After random assignment, the study imposed additional inclusion criteria related to participant effort, to ensure high quality responses. A common concern with online surveys is that some participants do not pay attention to instructions or read questions and just fill in responses haphazardly. The study built in several additional features to the survey to detect and remove potential participants demonstrating this type of behavior. These features include the following:

- **Dominated school.** The school ranking exercise included one school (named "Adams Elementary") that received the lowest rating by far on every measure in the display but was listed first alphabetically in the rank-ordering exercise. In other words, regardless of how schools were ordered in the information display, participants were always asked to rank schools by marking their selections on a list that appeared in alphabetical order, with the dominated school at the top. The study assumed anyone who selected this school was exhibiting low effort, in which case they were not included in the analysis. After removing those screened out for other reasons, the study removed 5 percent of potential participants from the remaining sample for selecting the dominated school.
- **Response time effort.** Following work by Wise and Kong (2005), the study screened out unmotivated potential participants by measuring the time to complete the survey from start to finish and time to complete selected items. The median completion time was 19 minutes among those who were not removed from the sample for other reasons. To determine a cutoff for reasonable response times, the research team conducted an analysis of the performance of rapid responders on the understanding module, because very low scores on the understanding module would indicate guessing. There was a steep drop-off in performance at the threshold of less than 7 minutes, compared to other cutoffs in completion times. It was assumed that participants who completed the entire survey in less than 7 minutes were rushing through and not answering questions sincerely. After removing potential participants screened out for other reasons, 1 percent of the sample was subsequently removed due to completion times of less than 7 minutes.

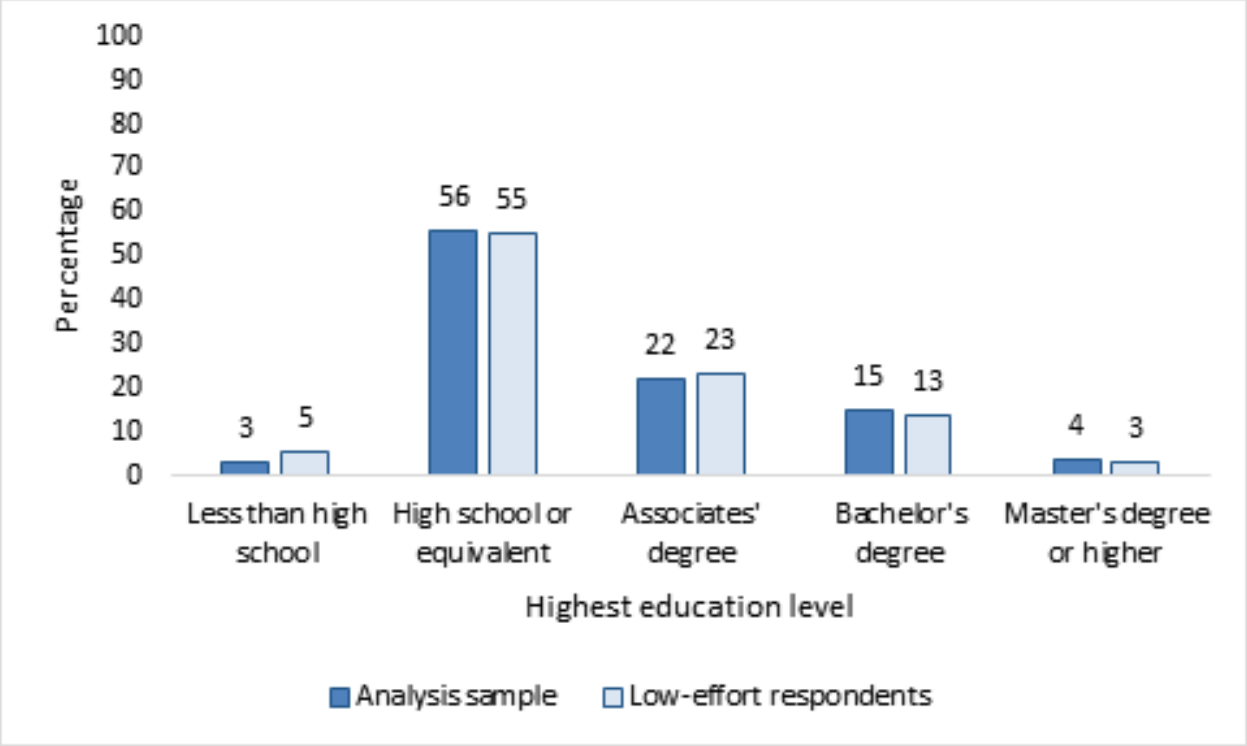
Table B.7. Flow of participants out of the sample

Status	Cases removed	Remaining participants
Started survey		6,160
Early dropout/screened out/non-consent		
Did not consent to participate	269	5,891
Screened out for demographic reasons (no children in age range, no role in education decisions of child, or income too high)	565	5,326
Quit before the IMC	700	4,626
Screened out due to incorrect responses on the IMC	494	4,132
Quit during the IMC but before randomization/exposure to treatment	171	3,961
Attrition		
Quit after randomization/exposure to treatment	228	3,736
Removed from analysis due to choosing Adams (dominated school)	182	3,551
Removed from analysis due to fast response time	51	3,500

Note: IMC = instructional manipulation check.

Sample members who were screened out because of low effort or inattentiveness did not differ in a statistically or substantively meaningful way from the analysis sample. Figure B.2 shows that their education levels were similar. A chi-square test of the independence of education and sample status could not be rejected. Similarly, there were negligible differences in the percentages who reported that the main language spoken at home was English (93 percent of the analysis sample and 91 percent of the low effort sample).

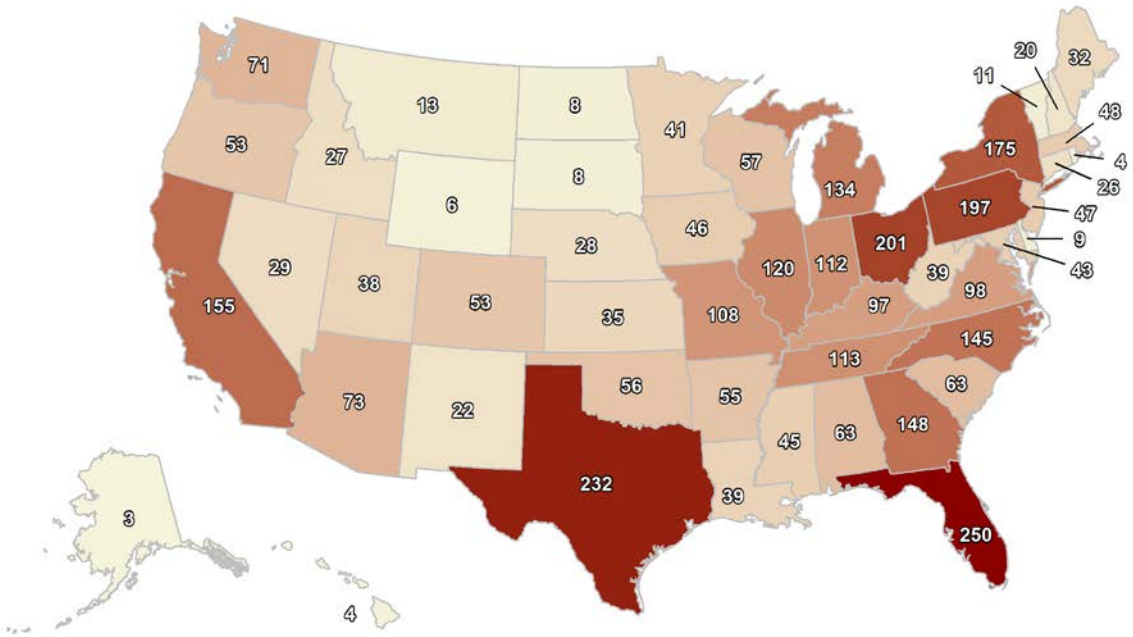
Figure B.2. Education level by screen-out status



b. Characteristics of the sample

A key limitation of the study is external validity, due to the use of an opt-in, non-probability sample. It is possible that those who sign up to participate in surveys such as this one may differ in distinct ways from the national population of low-income parents involved in making school choices. For example, those participating in online surveys may have more experience in viewing online information displays than other parents due to their time spent completing surveys. However, because the population of interest has not been studied previously, there is no way to compare the characteristics of the sample to those in the national population. Despite this limitation, the sample was geographically diverse, representing all 50 states (Figure B.3) and nearly evenly divided among urban, suburban, and rural participants (Table B.8).

Figure B.3. Number of participants from each state



Note: The map represents the analytic sample of 3,500 parents.

As intended, the sample included only low-income parents of school-aged children (Table B.8). About three-fourths (76 percent) of study participants were women. It is also worth noting that the vast majority (93 percent) of participants spoke only English at home. The survey was administered in English, so the study recruited only those who could complete an English-language questionnaire. Table B.8 presents information on the marital status and role of the respondents in making education decisions for their children. There is no way to ascertain whether study participants sought the help of a spouse or other family member in completing the survey, but the exercise was designed to be completed individually.

The sample members were not entirely new to school choice. About half of the sample members (52 percent) had public school options available to them in addition to neighborhood schools. These options can include charter schools, magnet schools, or open enrollment plans

that allow selection of different schools in the same district. Nearly two thirds (64 percent) of the sample had very low annual income, defined as less than $30,000. These characteristics – prior exposure to choice and very low income – overlapped with urbanicity. Specifically, nearly one-third of the sample (32 percent) were from urban areas and chi-square tests of the independence of prior school choice exposure and urbanicity and of very low income and urbanicity were both rejected at the 0.001 level. In urban areas, 69 percent of sample members were very low income and 59 percent had prior exposure to school choice (see Table B.9).

Table B.8. Demographic characteristics of the analytic sample

	Number	Percentage
Parent characteristics		
Female	2,642	75.5
Race/ethnicity		
White, non-Hispanic	2,482	70.9
Black, non-Hispanic	422	12.1
Hispanic	358	10.2
Other, non-Hispanic	238	6.8
Age		
Under 25	233	6.7
25–34 years	1,276	36.5
35–44 years	1,055	30.1
45 or older	936	26.7
Education		
Less than high school	113	3.2
High school diploma or GED	1,955	55.9
Associate's degree	776	22.2
Bachelor's degree or higher	656	18.7
Language spoken at home		
English only	3,254	93.0
Spanish	185	5.3
Other	61	1.7
Marital status		
Married or living with a partner	2,293	65.5
Respondent's role in child's education decisions		
Only person who makes decisions	1,214	34.7
The main person, but takes into account the opinion of the child or another adult	1,058	30.2
Share equally in the decision with the child or other adult	1,133	32.4
Involved, but in some other way	95	2.7
Household income, annual		
$10,000 or less	351	10.0
$10,001–$20,000	696	19.9
$20,001–$30,000	1,194	34.1
$30,001–$40,000	1,259	36.0
More than $40,000	0	0.0
Internet usage per week		
Less than 10 hours	786	22.5
10–29 hours	1,712	48.9
30 or more hours	974	27.8
Characteristics of parent's youngest child		
Child is female	1,677	47.9
Child has ever had an individualized education plan (IEP)	805	23.0
Community characteristics		
Public school options available in the community besides the neighborhood school (e.g., magnet, charter)	1,819	52.0
Urbanicity		
Urban	1,114	31.8
Suburban	1,307	37.3
Rural	1,079	30.8

Source: Parent information and school choice survey administered in August–October 2016.

Note: These demographics are for the analytic sample of 3,500 parents.

Table B.9. Income and prior choice exposure of urban and non-urban sample members

Characteristic	Urban	Non-Urban	Difference
Annual income			
Less than $30,000	69.2	61.7	7.5*
$30,000 to $40,000	30.9	38.4	-7.5
Prior choice exposure			
Yes	58.8	48.8	10.0*
No	41.2	51.2	-10.0

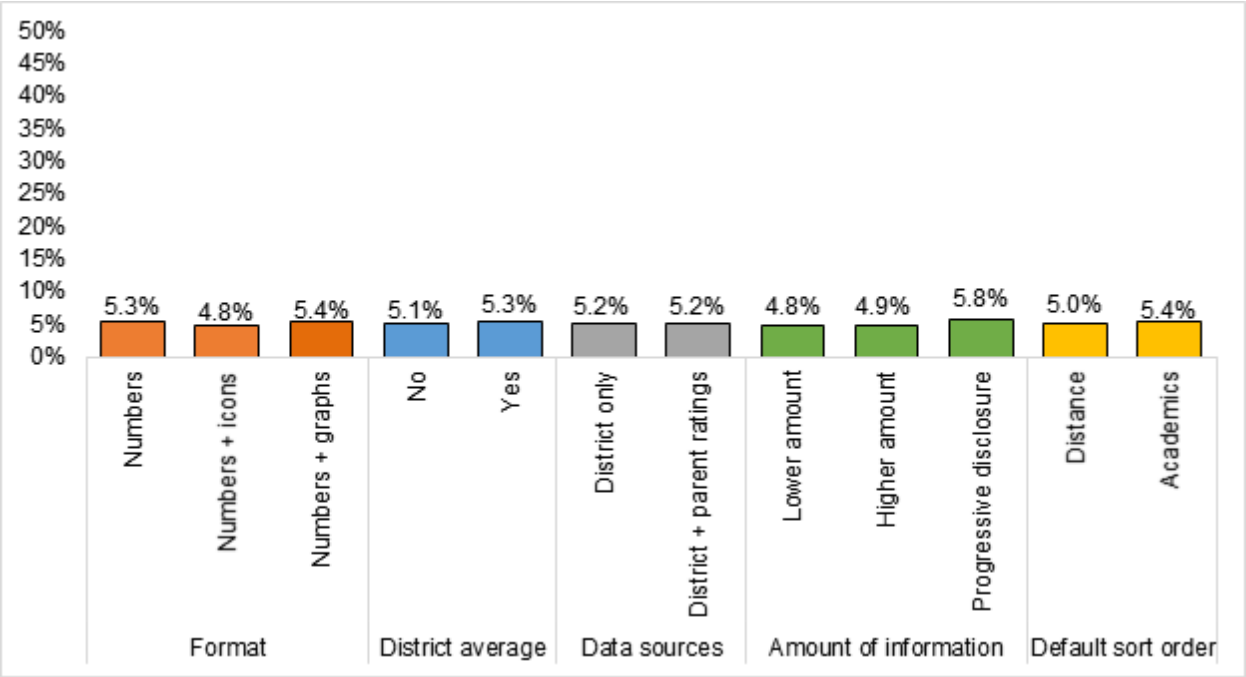
* Chi-square test of independence rejected at the 0.05 level.

c. Results from randomization

For an experiment to produce results that are not influenced by the characteristics of study participants, randomly assigned groups must be equivalent to each other after accounting for attrition. In this study, attrition was defined as participants having been randomly assigned, but then either not completing the outcomes survey or completing it with low effort or insincere effort as determined by the low-effort screening criteria described above. The overall attrition rate in the sample was less than 6 percent and did not differ meaningfully by display strategy (Figure B.4).

As expected, the groups randomly assigned to each display strategy in the study were equivalent at baseline (Tables B.10–B.12 show this for the analysis sample). Chi-square tests of independence of strategies (by factor) and baseline characteristics fail to reject the null hypothesis of "no baseline differences" in 53 out of the 55 tests the study conducted.

Figure B.4. Attrition, by display strategy



Note: To calculate attrition overall and for each display strategy, the total number included in the analysis were divided by the number of people randomly assigned, and this ratio was subtracted from 1. Overall attrition for the entire eligible sample was 5.5 percent. Differential attrition was less than 1 percentage point within each factor. Based on What Works Clearinghouse standards, overall attrition would have to be higher than 50 percent for 1 percentage point of differential attrition to create a risk of unacceptable bias (U.S. Department of Education 2014).

Table B.10. Parent demographic characteristics, by display strategy

Characteristic	Format			District average shown		Source of information		Amount of information			Default sort order	
	Numb.	Icon	Graph	No	Yes	District	Both	Low	High	Prog.	Dist.	Acad.
Female	74.3	76.0	76.2	74.6	76.4	75.6	75.4	74.6	74.6	77.3	75.4	75.6
Race/ethnicity												
White, non-Hispanic	69.3	71.4	72.1	70.3	71.5	70.3	71.5	70.6	71.0	71.1	71.2	70.6
Black, non-Hispanic	12.2	12.7	11.3	12.2	11.9	12.1	12.1	12.0	12.1	12.1	12.6	11.5
Hispanic	10.7	9.3	10.6	10.8	9.6	10.3	10.1	9.4	11.0	10.3	9.7	10.8
Other, non-Hispanic	7.8	6.6	6.0	6.6	7.0	7.3	6.3	8.0	5.9	6.5	6.5	7.1
Age												
Under 25	6.7	6.9	6.3	7.3	6.0	6.0	7.3	6.2	7.0	6.8	6.8	6.5
25–34 years	35.3	37.5	36.6	36.4	36.5	35.8	37.1	34.7	38.1	36.5	36.5	36.4
35 or older	58.0	55.6	57.1	56.3	57.5	58.2	55.6	59.1	54.9	56.7	56.7	57.1
Education												
Less than high school	3.3	3.3	3.0	3.2	3.3	3.6	2.9	2.8	3.1	3.7	3.2	3.2
High school diploma or GED	54.6	56.5	56.5	55.6	56.1	54.8	56.9	54.6	56.4	56.5	55.3	56.5
Associate's degree	23.4	21.2	21.9	22.1	22.3	22.0	22.3	23.2	22.9	20.5	23.8	20.6
Bachelor's degree or higher	18.6	19.0	18.6	19.2	18.3	19.6	17.9	19.3	17.6	19.3	17.7	19.8
Language spoken at home												
English only	93.0	93.2	92.7	92.3	93.7	92.9	93.0	94.4	92.2	92.4	92.9	93.0
Spanish	5.0	4.9	6.0	5.9	4.7	5.5	5.1	4.3	6.1	5.4	5.1	5.5
Other	2.0	2.0	1.3	1.8	1.7	1.6	1.9	1.3	1.7	2.2	2.0	1.5
Household income												
$10,000 or less	9.1	11.4	9.6	11.7^	8.4^	10.9	9.2	10.7	10.4	8.9	10.0	10.0
$10,001–$20,000	19.3	19.5	20.9	18.2^	21.6^	19.8	20.0	19.0	19.2	21.5	20.3	19.5
$20,001–$30,000	35.7	33.3	33.3	33.3^	34.9^	33.1	35.2	35.1	34.1	33.2	33.2	35.0
$30,001–$40,000	35.8	35.8	36.3	36.8^	35.1^	36.3	35.7	35.2	36.3	36.4	36.5	35.5
Internet usage per week												
Less than 10 hours	19.6^	25.3^	23.0^	23.1	22.1	22.1	23.1	22.4	23.1	22.4	22.9	22.3
10–29 hours	51.5^	47.7^	48.7^	48.0	50.6	50.7	48.0	50.1	48.7	49.1	50.0	48.7
30 or more hours	28.9^	27.0^	28.3^	28.9	27.3	27.2	28.9	27.5	28.2	28.5	27.1	29.0
Sample size (number of participants)	**1,175**	**1,169**	**1,156**	**1,764**	**1,736**	**1,742**	**1,758**	**1,164**	**1,159**	**1,177**	**1,743**	**1,757**

Source: Parent information and school choice survey administered in Aug.–Oct. 2016.

Note: Prog. = progressive disclosure; District = district source only; Both = district source and parent ratings; Numb. = number format; Acad. = academic sort order by default; Dist. = distance sort order by default.

Shaded blocks of cells, also denoted with a caret (^) are those for which a chi-square test of independence of rows and columns was rejected at the 0.05 level.

Table B.11. Demographic characteristics of parent's youngest child, by display strategy

	Format			District average shown		Source of information		Amount of information			Default sort order	
Characteristic	**Numb.**	**Icon**	**Graph**	**No**	**Yes**	**District**	**Both**	**Low**	**High**	**Prog.**	**Dist.**	**Acad.**
Child is female	49.9	45.9	47.9	47.7	48.1	48.3	47.5	47.6	49.1	47.1	47.6	48.2
Child has ever had an individualized education plan (IEP)	22.2	23.7	23.1	22.1	24.0	22.9	23.1	23.2	22.3	23.5	23.9	22.1
Sample size (participants)	**1,175**	**1,169**	**1,156**	**1,764**	**1,736**	**1,742**	**1,758**	**1,164**	**1,159**	**1,177**	**1,743**	**1,757**

Source: Parent information and school choice survey administered in Aug.–Oct. 2016.

Note: Prog. = progressive disclosure; District = district source only; Both = district source and parent ratings; Numb. = number format; Acad. = academic sort order by default; Dist. = distance sort order by default.
Chi-square tests of independence were conducted for each characteristic/factor combination. None was rejected at the 0.05 level.

Table B.12. Community characteristics of parents, by display strategy

	Format			District average shown		Source of information		Amount of information			Default sort order	
Characteristic	**Numb.**	**Icon**	**Graph**	**No**	**Yes**	**District**	**Both**	**Low**	**High**	**Prog.**	**Dist.**	**Acad.**
Public school options available in the community besides the neighborhood school (e.g., magnet, charter)	53.1	52.6	50.2	51.5	52.4	51.4	52.5	53.2	50.3	52.4	52.0	51.9
Urbanicity												
Urban	31.5	31.9	32.1	32.2	31.5	32.0	31.7	30.8	32.3	32.4	32.0	31.7
Suburban	37.3	36.9	37.9	37.7	37.0	36.5	38.2	36.7	37.8	37.6	37.7	37.0
Rural	31.2	31.2	30.0	30.1	31.6	31.6	30.1	32.5	29.9	30.1	30.4	31.3
Sample size (participants)	**1,175**	**1,169**	**1,156**	**1,764**	**1,736**	**1,742**	**1,758**	**1,164**	**1,159**	**1,177**	**1,743**	**1,757**

Source: Parent information and school choice survey administered in Aug.–Oct. 2016.

Note: Prog. = progressive disclosure; District = district source only; Both = district source and parent ratings; Numb. = number format; Acad. = academic sort order by default; Dist. = distance sort order by default.
Chi-square tests of independence were conducted for each characteristic/factor combination. None was rejected at the 0.05 level.

C. RESULTS

1. Impact results

This section reports the full set of impact estimates from the study, and presents additional sets of impact estimates for subsamples of parents that may be of interest to policymakers. The results of the experiment include the main effects of each display strategy on the full population of survey participants, main effects on relevant subsamples, and two-way interaction effects between strategies.

a. Main effects

Tables C.1 and C.2 present the main effects for the full sample. For each display strategy, Table C.1 shows the effect size (the impact in standard deviation units relative to the overall sample average) and the posterior probability of that strategy being the best-tested option for that factor with respect to the understanding, ease of use, and satisfaction outcomes. Table C.2 shows the effect sizes for each strategy on the characteristics of schools selected by parents, with separate sets of effects estimated for four different categories of school information (academics, distance, safety, and resources). Effects for each school information category were estimated in a separate regression model. As discussed in Section A.3 of this appendix, the hypothetical schools presented to parents in the study exemplified tradeoffs between these information categories: thus, if a strategy had a positive effect on selected schools in one information category, that strategy was likely to have a corresponding negative effect in one or more of the other three information categories.

The results in Tables C.1 and C.2 provide the same results presented in the guide; however, those in the guide are predicted means expressed in natural units, such as the percentage of points earned on the understanding exercise or the average distance in miles from home of selected schools. The guide used natural units to make findings more interpretable. This appendix, in contrast, uses effect sizes to facilitate comparisons across many outcomes and analyses. To calculate the predicted means shown in the guide, each effect size was multiplied by the standard deviation of the outcome and added to the mean of the outcome. For example, the experiment estimated that including parent survey ratings had an impact of 0.06 standard deviations on the satisfaction outcome. The satisfaction outcome, on a 0-100 scale, had a mean of 90 and a standard deviation of 21. The predicted mean reported in the guide in this case was 91.5, which is 90 plus the product of 0.06 times 21 (after rounding the result to the nearest 0.5). To provide a reference for making these conversions, the mean and standard deviation of each outcome is summarized below:

- *Understanding* (0-100 scale)*:* mean of 73, standard deviation of 30
- *Ease of Use* (0-100 scale): mean of 86, standard deviation of 23
- *Satisfaction* (0-100 scale): mean of 90, standard deviation of 21
- *School academics* (percent proficient on state test, 0-100 scale): mean of 65, standard deviation of 23
- *School distance from home* (miles): mean of 2, standard deviation of 1.4

- *School safety* (percent with no suspensions, 0-100 scale): mean of 92, standard deviation of 4.8
- *School resources* (number of laptops per 100 students): mean of 55, standard deviation of 20

The pattern of main effects observed in the study was not sensitive to pairwise interaction effects between strategies. Specifically, the conclusions about which strategies were best for each outcome (based on the study's 70 percent threshold) did not change when combinations of strategies were paired together in all possible ways.

While predicted mean effects of each display strategy may not change knowledge or attitudes by more than a few points on a 100-point scale, the cumulative effects of making a decision on each of the five factors simultaneously can be larger. In total, the study examined 72 different information displays. Table C.3 presents the predicted differences between the display that maximizes each outcome and the display that minimizes each outcome. Figures C.1, C.2, and C.3 show the displays that were predicted to maximize each outcome.

Table C.4 presents the same information, but for the effects on choices outcome. In that set of results, it is apparent that certain changes in combinations of display strategies can result in parents choosing schools with academic performance 19 percentile points higher than the performance of schools they might otherwise have chosen. The display that leads parents to choose to the greatest degree on the basis of academic performance is shown in Figure C.4.

b. Subsamples

The study measured the effects of each display strategy for the following four subsamples of interest (defined in Section B.1), including only parents with: school choice exposure, lower education levels, lower income levels, and less intensive levels of Internet use. For each subsample, the results were compared with the results for the overall sample using the criteria described in Section B.1 (checking for "overturned non-recommendations" or "different best strategies") to conclude whether the findings are inconsistent. As shown in Table C.5, most of the findings presented in the guide remained consistent if the sample is restricted to each of these subsamples. Of the 18 separate findings presented in the guide, 16 remained consistent for parents with lower incomes, 13 remained consistent for parents with the lower education, 17 remained consistent for parents with non-intensive Internet use, and 15 remained consistent for parents with prior experience with school choice. More specifically, the subsample results are as follows:

- *Format*: For understanding and ease of use, the results of all four subsamples were consistent with the overall sample of parents in the study. For satisfaction, in three subsamples the displays that included graphs were *not* more satisfying than the other displays: the inconsistency was observed among parents with prior school choice experience, parents with lower education levels, and parents with lower incomes. This differs from the full sample, where the graph format was more satisfying than other formats.
- *District Average*: For ease of use and satisfaction, the results were consistent for all four subsamples of parents. However, for understanding, the displays without district

averages were best for two groups: parents with prior school choice experience and parents with lower education levels. This differs from the full sample, where there was no difference in understanding.

- *Source of information*: For understanding and satisfaction, the results were consistent for all four subsamples of parents. However, for ease of use, the district-only approach was best for two groups: parents with prior school choice experience and parents with lower education levels. This differs from the full sample, where there was no difference in ease of use.
- *Amount of information*: For ease of use and satisfaction, the findings were consistent for all four subsamples of parents. However, for understanding, the displays with progressive disclosure were best for one group: parents with lower incomes. This differs from the full sample, where there was no difference in understanding for various amounts of information.
- *Default sort order*: For understanding, the results were consistent for all four subsamples of parents. However, for ease of use, sorting by academics was best for two groups: parents with lower education levels and parents who use the Internet less than 30 hours per week. This differs from the full sample, where there was no difference in ease of use. In addition, for satisfaction, sorting by academics was best for one group: parents with lower education levels. This differs from the full sample, where sorting by distance was more satisfying.
- *Effects on choices:* The pattern of effects on choices was consistent across all four subsamples. For all subsamples, format, amount of information, and default order produced a consistent pattern of effects on the academic quality of schools selected by parents.

The full set of results for each subsample are presented in the same way as those for the full sample, in Tables C.6–C.13.

Table C.1. Impacts on understanding, ease of use, and satisfaction, for the full sample

		Effect size			Probability of being the best strategy		
Factor	Strategy	Understanding	Ease of use	Satisfaction	Understanding	Ease of use	Satisfaction
Format	Numbers only	**0.06^**	0.00	0.00	**0.99^**	0.23	0.26
	Numbers + icons	-0.03	0.00	-0.03	0.01	0.43	0.03
	Numbers + graphs	-0.03	0.00	**0.03^**	0.00	0.34	**0.71^**
District average shown	No	0.00	**0.01^**	**0.02^**	0.52	**0.84^**	**0.86^**
	Yes	0.00	-0.01	-0.02	0.48	0.17	0.14
Source of information	District only	**0.01^**	0.01	-0.06	**0.84^**	0.68	0.00
	District + parent ratings	-0.01	-0.01	**0.06^**	0.16	0.32	**1.00^**
Amount of information	Lower amount	0.00	0.02	-0.02	0.32	0.66	0.04
	Higher amount	0.00	-0.01	**0.04^**	0.26	0.11	**0.91^**
	Progressive disclosure	0.00	0.00	-0.02	0.42	0.23	0.06
Default sort order	By distance	**0.02^**	0.00	**0.02^**	**0.89^**	0.50	**0.87^**
	By academics	-0.02	0.00	-0.02	0.11	0.50	0.13

Note: The first three columns of numbers report the impact of each display strategy on the relevant outcome in effect size (standard deviation) units, with effects estimated relative to the average outcome across all combinations of strategies in the experiment. The second set of three columns reports the probability, calculated from the Bayesian posterior distribution, that a given strategy outperforms the other tested strategies for that factor. Within a factor, these estimated probabilities always sum to 1 for a given outcome. Posterior probabilities above 0.70, and the corresponding effect sizes, are indicated with a caret (^) and bold blue text.

Table C.2. Impacts on choices (characteristics of schools selected by parents), for the full sample

Factor	Strategy	Effect size for Academics	Distance	Resources	Safety
Format	Numbers only	-0.05	0.01	0.02	-0.02
	Numbers + icons	**0.05^**	-0.03	-0.04	**0.12^**
	Numbers + graphs	0.00	0.02	0.02	-0.11
District average shown	No	-0.02	-0.02	0.01	-0.03
	Yes	0.02	0.02	-0.01	**0.03^**
Source of information	District only	-0.01	-0.01	**0.06^**	-0.06
	District + parent surveys	0.01	0.01	-0.06	**0.06^**
Amount of information	Lower amount	**0.04^**	0.00	-0.09	0.00
	Higher amount	-0.09	0.02	**0.18^**	-0.01
	Progressive disclosure	**0.05^**	-0.03	-0.09	0.01
Default sort order	By distance	-0.10	**0.19^**	-0.01	-0.01
	By academics	**0.10^**	-0.19	0.01	0.01

Note: Each column of the table summarizes the results of a separate regression. The effect sizes represent, in standard deviation units, the effect of each display strategy on the average z-score of selected schools within a given category of information. The caret (^) and bold blue text indicate when a strategy is likely to have impacted selections: effect sizes are highlighted when there is a probability greater than 0.70 (calculated from a Bayesian posterior distribution) that the strategy had a true effect greater than zero.

Table C.3. Predicted knowledge and attitudinal outcomes for best and worst factor combinations

Outcome	Factor Combination					*Predicted mean*	*Best-Worst*
	Source: Includes Parent Survey	Default Sort	District Reference	Format	Amount of Information		
Understandability (mean on 0-100 scale) [a]							
Best display	No	Distance	No	Numbers	Progressive Disclosure	76.3	5.1
Worst display	Yes	Academics	Yes	Graphs	High	71.1	
Ease of Use (mean on 0-100 scale) [b]							
Best display	No	Academics	No	Icons	Low	86.7	1.2
Worst display	Yes	Distance	Yes	Numbers	High	85.5	
Satisfaction (mean on 0-100 scale) [c]							
Best display	Yes	Distance	No	Graphs	High	93.3	6.2
Worst display	No	Academics	Yes	Icons	Low	87.1	

[a] An understandability score of 73.3 (the study sample average) means that respondents correctly answered 73.3% of the factual questions about school attributes.

[b] An ease-of-use score of 86 (the study sample average) means that respondents "agree" or "strongly agree" with 86.1% of the questions about how easy the information was to use.

[c] An satisfaction score of 90 (the study sample average) means that respondents "agree" or "strongly agree" with 90% of the questions about how satisfying the information was to use.

Figure C.1. Most understandable display

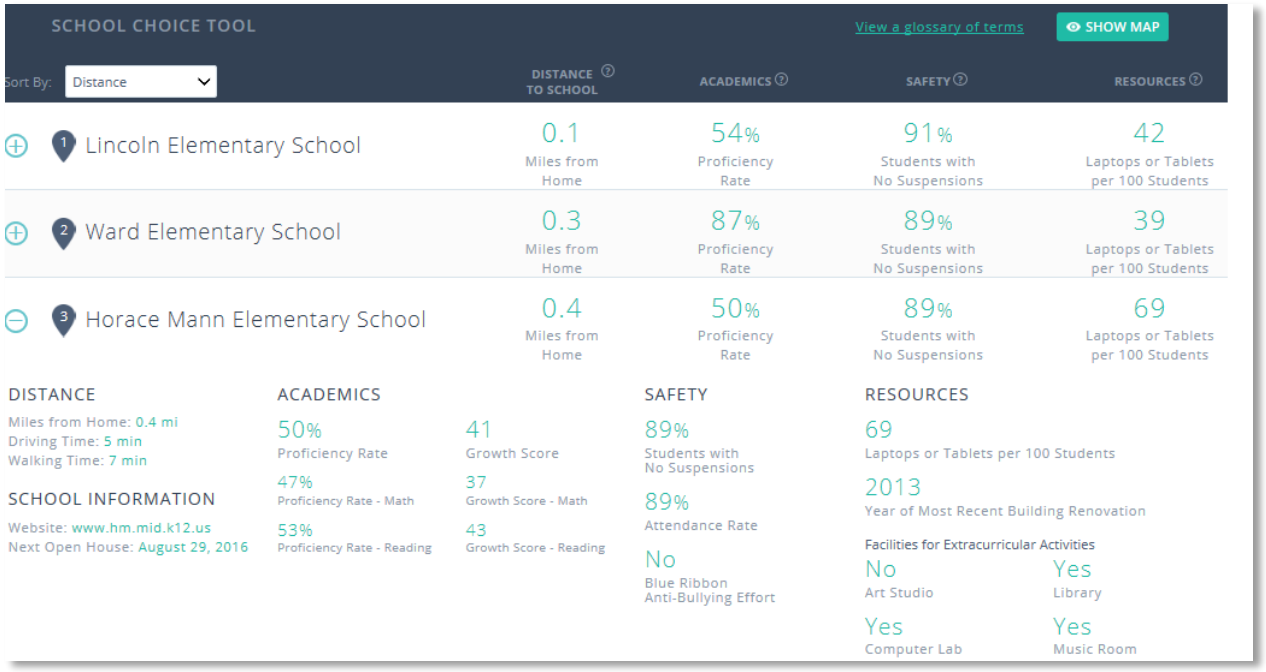


Figure C.2. Display with the greatest predicted ease of use

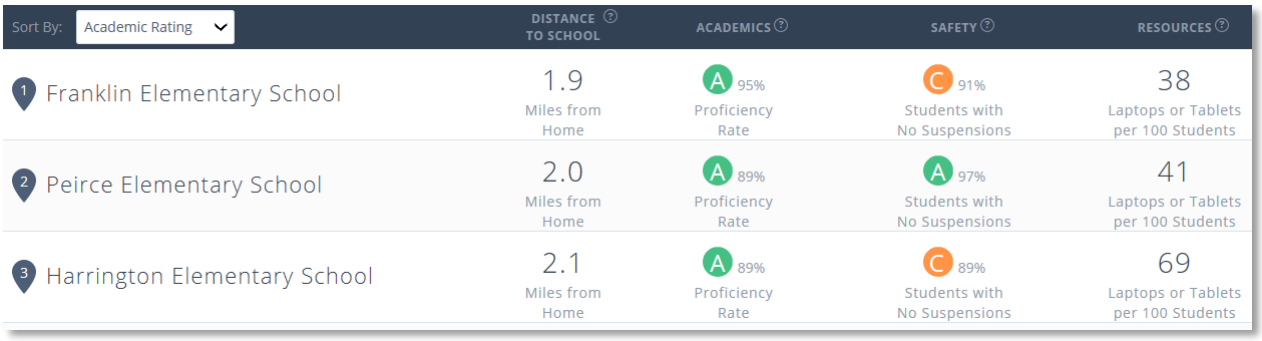


Figure C.3. Display with the greatest predicted level of satisfaction

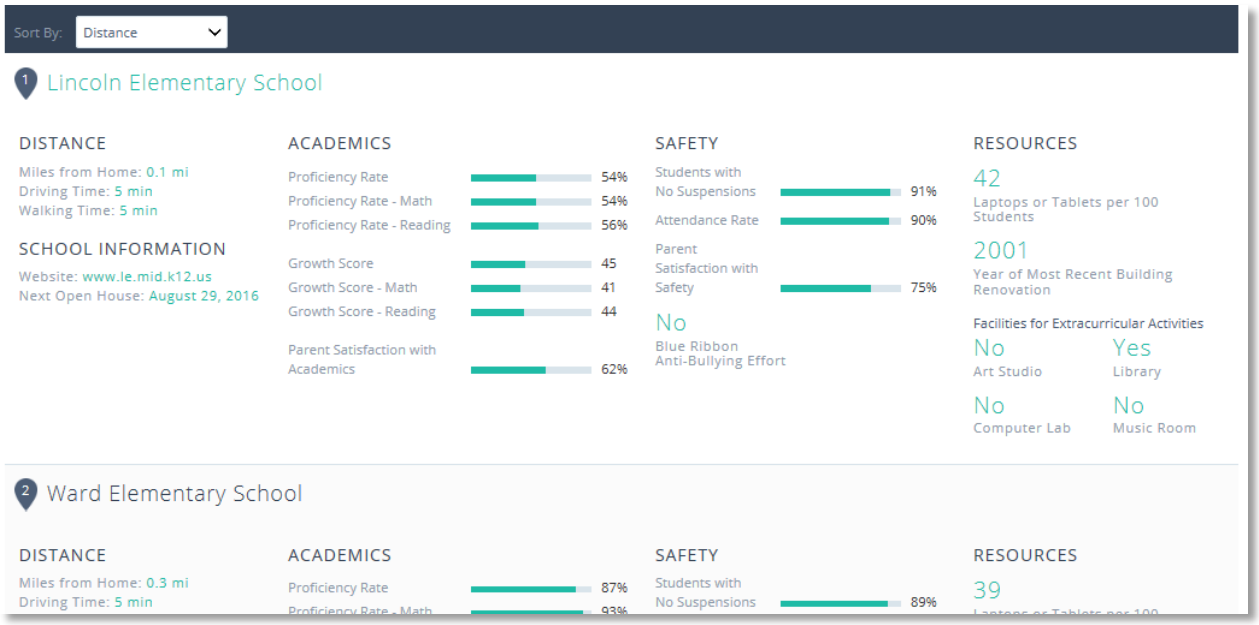


Table C.4. Predicted effects on choice for best and worst factor combinations

Outcome	Factor Combination					*Predicted mean*	*Best-Worst*
	Source: Includes Parent Survey	Default Sort	District Reference	Format	Amount of Information		
Academic Performance of Chosen Schools (percentile)							
Best display	Yes	Academics	Yes	Icons	Low	58.7	19.3
Worst display	No	Distance	No	Numbers	High	39.4	
Proximity of Chosen Schools (percentile)							
Best display	Yes	Distance	No	Graphs	High	60.3	20.9
Worst display	No	Academics	Yes	Icons	Progressive disclosure	39.4	
Safety of Chosen Schools (percentile)							
Best display	Yes	Academics	Yes	Icons	Progressive disclosure	59.1	17.0
Worst display	No	Distance	No	Graphs	High	42.1	
Resources of Chosen Schools (percentile)							
Best display	No	Academics	No	Numbers	High	60.6	18.6
Worst display	Yes	Distance	Yes	Icons	Low	42.1	

Note: A percentile score of 50 represents the school with the median value of the attribute among the 16 schools in the hypothetical district used for the study.

Figure C.4. Display that leads parents to select the highest academically-rated schools

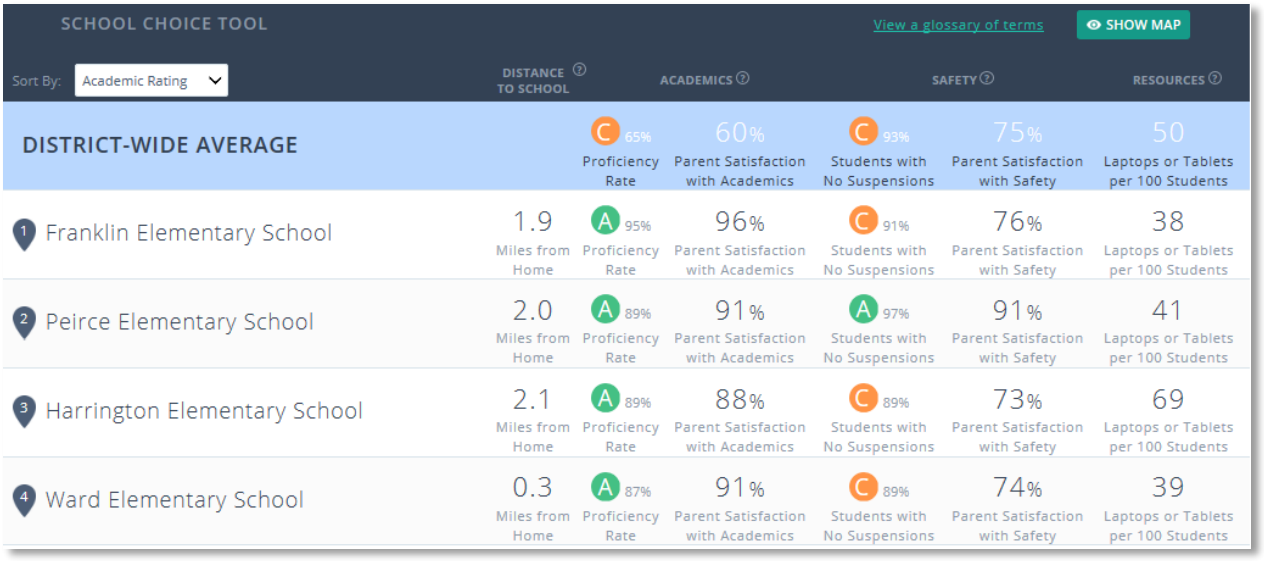


Table C.5. Comparison of subsample results to results for the full sample

Research question	Outcome	Best Strategy (Largest Impact, Posterior Probability >70%)				
			Subsample			
		Full Sample (N=3,500)	Prior school choice experience (N=1,819)	No college completion (N=2,068)	Income < $30,000 per year (N=2,241)	Internet use < 30 hours per week (N=2,498)
1. Format	Understanding	**Numbers only**	*C*	*C*	*C*	*C*
	Ease of use	**No differences**	*C*	*C*	*C*	*C*
	Satisfaction	**Numbers + graphs**	*Numbers + graphs not best strategy*[a]	*Numbers + graphs not best strategy*[a]	*Numbers + graphs not best strategy*[a]	*C*
2. District average shown	Understanding	**No differences**	*Without district average is best*	*Without district average is best*	*C*	*C*
	Ease of use	**Without district average**	*C*	*C*	*C*	*C*
	Satisfaction	**Without district average**	*C*	*C*	*C*	*C*
3. Source of information	Understanding	**District only**	*C*	*C*	*C*	*C*
	Ease of use	**No differences**	*District only is best*	*District only is best*	*C*	*C*
	Satisfaction	**District + parent ratings**	*C*	*C*	*C*	*C*
4. Amount of information	Understanding	**No differences**	*C*	*C*	*Progressive disclosure is best*	*C*
	Ease of use	**No differences**	*C*	*C*	*C*	*C*
	Satisfaction	**Higher amount**	*C*	*C*	*C*	*C*
5. Default sort order	Understanding	**By distance**	*C*	*C*	*C*	*C*
	Ease of use	**No differences**	*C*	*By academics is best*	*C*	*By academics is best*
	Satisfaction	**By distance**	*C*	*By academics is best*	*C*	*C*
6. Effects on choices	Encouraging academics	**Sort by academics**	*C*	*C*	*C*	*C*
	Encouraging academics	**Format with numbers + icons**	*C*	*C*	*C*	*C*
	Encouraging academics	**Lower amount of information or progressive disclosure**	*C*	*C*	*C*	*C*

Note: *C* Subsample finding is consistent with full sample finding.

[a] The format with the largest effect size for this outcome in this subsample (numbers only) differed from the overall sample. The probability that the numbers-only format was the best did not exceed 70 percent, but the subsample finding is inconsistent with the full sample finding by the "different best strategy" rule discussed in Section B.1.

Table C.6. Impacts on understanding, ease of use, and satisfaction, for parents with school choice exposure

		Effect size			Probability of being the best strategy		
Factor	Strategy	Understanding	Ease of use	Satisfaction	Understanding	Ease of use	Satisfaction
Format	Numbers only	**0.06^**	0.01	0.02	**0.89^**	0.42	0.57
	Numbers + icons	0.00	0.01	-0.01	0.10	0.50	0.16
	Numbers + graphs	-0.07	-0.02	0.00	0.00	0.08	0.28
District average shown	No	**0.01^**	0.01	**0.01^**	**0.75^**	0.65	**0.70^**
	Yes	-0.01	-0.01	-0.01	0.25	0.35	0.30
Source of information	District only	**0.02^**	**0.02^**	-0.03	**0.86^**	**0.83^**	0.07
	District + parent ratings	-0.02	-0.02	**0.03^**	0.14	0.17	**0.93^**
Amount of information	Lower amount	-0.01	0.00	-0.01	0.23	0.26	0.20
	Higher amount	0.01	0.00	0.02	0.51	0.36	0.62
	Progressive disclosure	-0.01	0.00	-0.01	0.26	0.38	0.18
Default sort order	By distance	0.00	0.00	**0.02^**	0.51	0.47	**0.84^**
	By academics	0.00	0.00	-0.02	0.49	0.53	0.16

Note: The first three columns of numbers report the impact of each display strategy on the relevant outcome in effect size (standard deviation) units, with effects estimated relative to the average outcome across all combinations of strategies in the experiment. The second set of three columns reports the probability, calculated from the Bayesian posterior distribution, that a given strategy outperforms the other tested levels for that factor. Within a factor, these estimated probabilities always sum to 1 for a given outcome. Posterior probabilities above 0.70, and the corresponding effect sizes, are indicated with a caret (^) and bold blue text.

Table C.7. Impacts on choices (characteristics of schools selected by parents), for parents with school choice exposure

Factor	Strategy	Effect size for Academics	Distance	Resources	Safety
Format	Numbers only	-0.03	0.01	0.01	0.00
	Numbers + icons	**0.04^**	-0.03	-0.03	**0.10^**
	Numbers + graphs	-0.01	0.02	0.02	-0.09
District average shown	No	0.00	-0.01	0.01	-0.01
	Yes	0.00	0.01	-0.01	0.01
Source of information	District only	0.01	0.00	0.02	-0.06
	District + parent surveys	-0.01	0.00	-0.02	**0.06^**
Amount of information	Lower amount	0.02	**0.04^**	-0.09	-0.01
	Higher amount	-0.07	0.01	**0.20^**	-0.02
	Progressive disclosure	**0.05^**	-0.05	-0.11	**0.03^**
Default sort order	By distance	-0.09	**0.20^**	0.00	-0.03
	By academics	**0.09^**	-0.20	0.00	**0.03^**

Note: Each column of the table summarizes the results of a separate regression. The effect sizes represent, in standard deviation units, the effect of each display strategy on the average z-score of selected schools within a given category of information. The bold blue text and caret (^) indicate when a strategy is likely to have impacted selections: effect sizes are highlighted when there is a probability greater than 0.70 (calculated from a Bayesian posterior distribution) that the strategy had a true effect greater than zero.

Table C.8. Impacts on understanding, ease of use, and satisfaction, for parents with lower education levels

		Effect size			Probability of being the best strategy		
Factor	Strategy	Understanding	Ease of use	Satisfaction	Understanding	Ease of use	Satisfaction
Format	Numbers only	**0.05^**	-0.01	0.02	**0.87^**	0.19	0.62
	Numbers + icons	-0.02	0.02	-0.03	0.07	0.68	0.05
	Numbers + graphs	-0.02	-0.01	0.01	0.06	0.13	0.33
District average shown	No	**0.02^**	**0.02^**	0.01	**0.81^**	**0.82^**	0.66
	Yes	-0.02	-0.02	-0.01	0.19	0.18	0.35
Source of information	District only	**0.04^**	0.01	-0.06	**0.96^**	0.67	0.00
	District + parent ratings	-0.04	-0.01	**0.06^**	0.04	0.33	**1.00^**
Amount of information	Lower amount	-0.02	0.02	-0.03	0.11	0.54	0.08
	Higher amount	0.00	-0.03	**0.03^**	0.21	0.08	**0.71^**
	Progressive disclosure	0.02	0.01	0.00	0.68	0.38	0.22
Default sort order	By distance	**0.01^**	-0.01	-0.02	**0.77^**	0.23	0.18
	By academics	-0.01	**0.01^**	**0.02^**	0.23	**0.77^**	**0.82^**

Note: The first three columns of numbers report the impact of each display strategy on the relevant outcome in effect size (standard deviation) units, with effects estimated relative to the average outcome across all combinations of strategies in the experiment. The second set of three columns reports the probability, calculated from the Bayesian posterior distribution, that a given strategy outperforms the other tested levels for that factor. Within a factor, these estimated probabilities always sum to 1 for a given outcome. Posterior probabilities above 0.70, and the corresponding effect sizes, are indicated with a caret (^) and bold blue text.

Table C.9. Impacts on choices (characteristics of schools selected by parents), for parents with lower education levels

Factor	Strategy	Effect size for Academics	Distance	Resources	Safety
Format	Numbers only	-0.09	**0.03^**	**0.05^**	-0.02
	Numbers + icons	**0.06^**	-0.03	-0.05	**0.12^**
	Numbers + graphs	**0.03^**	0.00	0.00	-0.10
District average shown	No	-0.01	0.01	-0.03	-0.01
	Yes	0.01	-0.01	**0.03^**	0.01
Source of information	District only	-0.01	0.00	**0.08^**	-0.05
	District + parent ratings	0.01	0.00	-0.08	**0.05^**
Amount of information	Lower amount	**0.07^**	0.00	-0.11	-0.01
	Higher amount	-0.10	0.02	**0.17^**	0.00
	Progressive disclosure	**0.03^**	-0.02	-0.06	0.01
Default sort order	By distance	-0.11	**0.19^**	0.02	-0.02
	By academics	**0.11^**	-0.19	-0.02	0.02

Note: Each column of the table summarizes the results of a separate regression. The effect sizes represent, in standard deviation units, the effect of each display strategy on the average z-score of selected schools within a given category of information. The bold blue text and caret (^) indicate when a strategy is likely to have impacted selections: effect sizes are highlighted when there is a probability greater than 0.70 (calculated from a Bayesian posterior distribution) that the strategy had a true effect greater than zero.

Table C.10. Impacts on understanding, ease of use, and satisfaction, for parents with lower income levels

		Effect size			Probability of being the best strategy		
Factor	Strategy	Understanding	Ease of use	Satisfaction	Understanding	Ease of use	Satisfaction
Format	Numbers only	**0.03^**	-0.01	0.02	**0.79^**	0.12	0.59
	Numbers + icons	-0.01	0.02	-0.03	0.12	0.60	0.05
	Numbers + graphs	-0.02	0.00	0.01	0.09	0.28	0.36
District average shown	No	0.00	**0.01^**	**0.01^**	0.58	**0.80^**	**0.77^**
	Yes	0.00	-0.01	-0.01	0.42	0.20	0.23
Source of information	District only	**0.03^**	**0.01^**	-0.04	**0.95^**	**0.72^**	0.03
	District + parent ratings	-0.03	-0.01	**0.04^**	0.05	0.28	**0.97^**
Amount of information	Lower amount	-0.01	0.02	-0.02	0.13	0.64	0.09
	Higher amount	-0.02	-0.02	0.02	0.07	0.10	0.64
	Progressive disclosure	**0.03^**	0.00	0.00	**0.80^**	0.26	0.27
Default sort order	By distance	**0.02^**	0.00	**0.02^**	**0.83^**	0.38	**0.90^**
	By academics	-0.02	0.00	-0.02	0.17	0.62	0.10

Note: The first three columns of numbers report the impact of each display strategy on the relevant outcome in effect size (standard deviation) units, with effects estimated relative to the average outcome across all combinations of strategies in the experiment. The second set of three columns reports the probability, calculated from the Bayesian posterior distribution, that a given strategy outperforms the other tested levels for that factor. Within a factor, these estimated probabilities always sum to 1 for a given outcome. Posterior probabilities above 0.70, and the corresponding effect sizes, are indicated with a caret (^) and bold blue text.

Table C.11. Impacts on choices (characteristics of schools selected by parents), for parents with lower income levels

Factor	Strategy	Effect size for Academics	Distance	Resources	Safety
Format	Numbers only	-0.07	0.01	**0.04^**	-0.03
	Numbers + icons	**0.05^**	-0.03	-0.03	**0.11^**
	Numbers + graphs	0.02	**0.03^**	-0.01	-0.09
District average shown	No	-0.01	0.01	0.00	-0.03
	Yes	0.01	-0.01	0.00	**0.03^**
Source of information	District only	-0.02	-0.01	**0.05^**	-0.04
	District + parent ratings	0.02	0.01	-0.05	**0.04^**
Amount of information	Lower amount	0.02	0.00	-0.08	0.00
	Higher amount	-0.07	**0.04^**	**0.16^**	-0.01
	Progressive disclosure	**0.04^**	-0.04	-0.08	0.01
Default sort order	By distance	-0.07	**0.16^**	-0.03	0.00
	By academics	**0.07^**	-0.16	**0.03^**	0.00

Note: Each column of the table summarizes the results of a separate regression. The effect sizes represent, in standard deviation units, the effect of each display strategy on the average z-score of selected schools within a given category of information. The bold blue text and caret (^) indicate when a strategy is likely to have impacted selections: effect sizes are highlighted when there is a probability greater than 0.70 (calculated from a Bayesian posterior distribution) that the strategy had a true effect greater than zero.

Table C.12. Impacts on understanding, ease of use, and satisfaction, for parents with non-intensive Internet use

		Effect size			Probability of being the best strategy		
Factor	Strategy	Understanding	Ease of use	Satisfaction	Understanding	Ease of use	Satisfaction
Format	Numbers only	**0.05^**	0.00	0.01	**0.94^**	0.32	0.32
	Numbers + icons	-0.02	0.01	-0.03	0.03	0.52	0.03
	Numbers + graphs	-0.03	-0.01	0.03	0.02	0.15	0.65
District average shown	No	0.00	**0.02^**	**0.02^**	0.54	**0.85^**	**0.86^**
	Yes	0.00	-0.02	-0.02	0.46	0.15	0.14
Source of information	District only	0.00	0.00	-0.05	0.46	0.58	0.00
	District + parent ratings	0.00	0.00	**0.05^**	0.54	0.42	**1.00^**
Amount of information	Lower amount	0.00	0.02	0.00	0.27	0.69	0.22
	Higher amount	-0.01	-0.02	0.02	0.14	0.11	0.69
	Progressive disclosure	0.01	0.00	-0.02	0.58	0.20	0.09
Default sort order	By distance	0.01	-0.01	**0.02^**	0.64	0.20	**0.82^**
	By academics	-0.01	**0.01^**	-0.02	0.36	**0.80^**	0.18

Note: The first three columns of numbers report the impact of each display strategy on the relevant outcome in effect size (standard deviation) units, with effects estimated relative to the average outcome across all combinations of strategies in the experiment. The second set of three columns reports the probability, calculated from the Bayesian posterior distribution, that a given strategy outperforms the other tested levels for that factor. Within a factor, these estimated probabilities always sum to 1 for a given outcome. Posterior probabilities above 0.70, and the corresponding effect sizes, are indicated with a caret (^) and bold blue text.

Table C.13. Impacts on choices (characteristics of schools selected by parents), for parents with non-intensive Internet use

		Effect size for			
		Academics	Distance	Resources	Safety
Format	Numbers only	-0.03	0.00	0.01	-0.03
	Numbers + icons	**0.04^**	-0.01	-0.04	**0.12^**
	Numbers + graphs	-0.01	0.01	**0.03^**	-0.09
District average shown	No	-0.01	0.01	0.01	-0.01
	Yes	0.01	-0.01	-0.01	0.01
Source of information	District only	-0.01	-0.01	**0.07^**	-0.08
	District + parent ratings	0.01	0.01	-0.07	**0.08^**
Amount of information	Lower amount	**0.02^**	0.01	-0.08	-0.02
	Higher amount	-0.10	**0.02^**	**0.17^**	0.01
	Progressive disclosure	**0.07^**	-0.04	-0.09	0.00
Default sort order	By distance	-0.09	**0.20^**	-0.01	-0.03
	By academics	**0.09^**	-0.20	0.01	**0.03^**

Note: Each column of the table summarizes the results of a separate regression. The effect sizes represent, in standard deviation units, the effect of each display strategy on the average z-score of selected schools within a given category of information. The bold blue text and caret (^) indicate when a strategy is likely to have impacted selections: effect sizes are highlighted when there is a probability greater than 0.70 (calculated from a Bayesian posterior distribution) that the strategy had a true effect greater than zero.

2. Sensitivity analyses

This subsection provides an overview of the sensitivity tests the study conducted, the rationale for each test, and a summary of the results.

a. Alternative outcome measures

Understanding. In addition to the primary understanding index, the study also tested an alternative understanding index that accounted for completion time. The study scored this alternative index by awarding additional bonus points for fast completion times. In this index, participants were eligible for a bonus for the item if they answered it correctly and had a speed in the top 25 percent of those who answered the question correctly. Constructing the outcome in this way tests whether each display strategy had an impact on a participant's likelihood of answering correctly with a speed in the top quartile for the sample. The bonus point awarded was worth half of the value of the item (such that the first four items could each result in a 0.5-point bonus, the fifth item in a 1-point bonus, and the final item in a 1.5-point bonus). Scores on this alternative measure ranged from 0.0 to 13.5.

The primary rationale for incorporating response time into the understanding scores is that processing speed may signal how well a person understands information. Also, the index with the time bonus shows greater variation across participants compared to the index without the time bonus. However, this experiment's time measurements may not always reflect true processing speed. For example, a participant could take a break in the middle of a question. It is also not clear that the amount of time needed to select the correct school is more related to understanding than it is to the usability of the profiles—a separate outcome measure. Another drawback of the index with the time bonus is that the units of the outcome measure cannot be interpreted easily. For these reasons, the study used the understanding index without the time bonus as the primary understanding measure but used the version with the time bonus as a sensitivity test.

Usability. As described in B.2, the study scored the ease of use and satisfaction measures using a binary 0–1 scale for each item, where 0 indicated strongly disagree or disagree and 1 indicated agree or strongly agree, and item scores were averaged to create an index ranging from 0 to 1. The primary rationale for this construction was that it could easily be interpreted as the proportion of items with which participants agreed. However, the study also tested a version of the index that used the full 1–4 scale of the items, treating a difference between "disagree strongly" and "disagree" equivalently to a difference between "disagree" and "agree." The advantage of the 1–4 scale is that it allows for greater variance; the disadvantage is that the units would be more difficult to interpret.

Selection of schools. In examining the effects of parents' school selections on the values they placed on school measures, the study gave greater weight to parents' top-ranked school than the second-ranked school, and greater weight to their second-ranked than third-ranked school. Specifically, the top-ranked school received double the weight of the second-ranked school, and the second-ranked school received double the weight of the third-ranked school. This approach assumes that the top-choice school is a better way to gauge parents' values than the third-ranked school. However, it is also possible that parents may not have differed strongly in their

preferences between their first and third choices. Therefore, the study also tested the selection of schools outcome where the measures of all three schools selected were weighted equally.

Sensitivity analysis results. For the understanding outcome, incorporating the time bonus into the understanding measure resulted in larger effect sizes for both amount of information and information sources (Table C.14). Parents did not perform as well on the understanding task when they had a larger amount of information compared to a lower amount or progressive disclosure (in which the lower amount of information was shown by default). Similarly, they performed worse on understanding when the study added parent ratings, compared to when they were presented only with the district data. Both results suggest that displays with more information take more time to review before finding the information needed to respond to a factual question about a school. All other effect sizes were similar between the original and time-bonus versions.

Using the four-point scale for the ease of use and satisfaction indexes resulted in effect sizes in the same direction and with similar magnitudes to the original binary versions of these scales. For both ease of use and satisfaction, the absolute differences between the original and the four-point versions ranged from 0.00 to 0.02.

The use of equal weighting of attributes for the three selected schools did not change the direction of any of the effects on school choices. For example, the two weighting formulas resulted in nearly the same effect sizes on choosing a higher academically performing school (Table C.14). The direction of effects also was consistent across the two weighting formulas for the selection of schools with closer distance and higher levels of safety and resources.

Table C.14. Comparison of effect sizes of original understanding, ease of use, satisfaction, and effects on choices outcomes to alternatives

		Understanding		Ease of use		Satisfaction		Choice based on academics	
		Original	Time bonus	Original	4-point scale	Original	4-point scale	Original	Equal weights
Format	Numbers only	0.06	0.06	0.00	0.00	0.00	-0.01	-0.05	-0.05
	Numbers + icons	-0.03	-0.01	0.00	0.01	-0.03	-0.01	0.00	-0.01
	Numbers + graphs	-0.03	-0.04	0.00	-0.01	0.03	0.03	0.05	0.06
District average shown	No	0.00	0.00	0.01	0.01	0.02	0.01	-0.02	-0.02
	Yes	0.00	0.00	-0.01	-0.01	-0.02	-0.01	0.02	0.02
Source of information	District only	0.01	0.04	0.01	0.01	-0.06	-0.05	-0.01	0.00
	District + parent ratings	-0.01	-0.04	-0.01	-0.01	0.06	0.05	0.01	0.00

		Understanding		Ease of use		Satisfaction		Choice based on academics	
		Original	Time bonus	Original	4-point scale	Original	4-point scale	Original	Equal weights
Amount of information	Lower amount	0.00	0.04	0.02	0.02	-0.02	-0.02	0.04	0.05
	Higher amount	0.00	-0.08	-0.01	-0.02	0.04	0.03	-0.09	-0.09
	Progressive disclosure	0.00	0.04	0.00	0.00	-0.02	-0.01	0.05	0.04
Default sort order	By distance	0.02	0.02	0.00	0.02	0.02	0.02	-0.10	-0.09
	By academics	-0.02	-0.02	0.00	-0.02	-0.02	-0.02	0.10	0.09

Note: The table reports the impact of each display strategy on the relevant outcome in effect size (standard deviation) units, with effects estimated relative to the average outcome across all combinations of strategies in the experiment.

b. Alternative model specifications and sample definitions

No covariates. To improve the precision of the estimates, the impact model includes several covariates—household income; parent's education; computer usage; experience with school choice; and characteristics of the youngest child, including gender, race/ethnicity, and whether they had received an individualized education program. The study also tested whether the results would hold if the model had no covariates. Because this study used random assignment with a simple random assignment algorithm (one in which all participants had the same likelihood of being assigned to each of the 72 treatment arms of the experiment), some researchers would argue that a simple model with no covariates provides a direct and appropriate way to obtain unbiased impact estimates. This sensitivity test examined whether the results are robust to this simpler model specification.

Frequentist analogue. The study's original protocol registered with the American Economic Association[2] and power calculations all assumed that the analysis would use Bayesian methods. The study conducted this sensitivity test to examine if there is any evidence that a frequentist analysis (that is, a model without hierarchical priors or Bayesian modeling) would have arrived at different conclusions, with the caveat that the study's sample size does not provide adequate statistical power for frequentist tests of statistical significance for small effect sizes.

Alternative sample definition. As described in Section B.3, the study eliminated from the sample those parents who selected Adams Elementary as one of their top three choices. The purpose of eliminating these participants was to remove those who were not making a sincere effort to do the survey. However, one could argue that choosing this school might indicate confusion about the information provided in the displays; this would be problematic for the study's internal validity if the propensity to select Adams Elementary was correlated with one of

[2] This study is registered in the AEA RCT Registry and the unique identifying number is: AEARCTR-0001190.

the display strategies tested in the experiment. For this reason, the study also included a sensitivity test that retained those participants who selected Adams.

Sensitivity analysis results. There were no meaningful differences between the magnitude and direction of effect sizes using the original model and sample, compared to the tested alternatives. On the understanding outcome, the absolute differences between the original and any of the alternatives ranged from 0.00 to 0.02 (Table C.15). Though not illustrated in the table below, these findings were consistent with the results of sensitivity tests with respect to the satisfaction and ease of use outcomes. For the "frequentist analog" sensitivity test, using a conventional significance test with a *p*-value of 0.05 did cause a substantial loss of precision. Only one of the effects (the effect of the numbers-only format on understanding) was statistically significant under this approach, even though all other point estimates remained similar in magnitude and sign to the estimates in the study's preferred Bayesian model.

Table C.15. Comparison of effect sizes for the understanding outcome to effect sizes with alternative model specifications and sample definitions

		Original	No covariates	Frequentist analogue	Sample with Adams choosers
Format	Numbers only	0.06	0.07	0.08*	0.05
	Numbers + icons	-0.03	-0.03	-0.03	-0.03
	Numbers + graphs	-0.03	-0.04	-0.05	-0.03
District average shown	No	0.00	0.00	0.00	0.01
	Yes	0.00	0.00	0.00	-0.01
Source of information	District only	0.01	0.01	0.02	0.01
	District + parent ratings	-0.01	-0.01	-0.02	-0.01
Amount of information	Lower amount	0.00	0.00	0.00	0.00
	Higher amount	0.00	0.00	-0.01	-0.01
	Progressive disclosure	0.00	0.00	0.01	0.01
Default sort order	By distance	0.02	0.02	0.02	0.01
	By academics	-0.02	-0.02	-0.02	-0.01

Note: The table reports the impact of each display strategy on the relevant outcome in effect size (standard deviation) units, with effects estimated relative to the average outcome across all combinations of strategies in the experiment.

* For the frequentist analogue sensitivity test, results were statistically significant at the 0.05 level with a Benajamini-Hochberg adjustment for multiple comparisons that accounts for the total number of display strategies (12) in the experiment.

This page has been left blank for double-sided copying.

REFERENCES

Akaah, I.P., and P.K. Korgaonkar. "An Empirical Comparison of the Predictive Validity of Self-Explicated, Huber-Hybrid, Traditional Conjoint, and Hybrid Conjoint Models." *Journal of Marketing Research*, vol. 20, no. 2, 1983, pp. 187–197.

Bell, Courtney A. "All Choices Created Equal? The Role of Choice Sets in the Selection of Schools." *Peabody Journal of Education,* vol. 84, no. 2, 2009, pp. 191–208.

Berger, James O. *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York: Springer-Verlag, 1985.

Borgmeier and Westenhoefer "Impact of different food label formats on healthiness evaluation and food choice of consumers: a randomized-controlled study." *BMC Public Health*, vol. 9, 2009, pp. 1-12.

Brooke, J. *System Usability Scale (SUS): A Quick-and-Dirty Method of System Evaluation User Information.* Reading, UK: Digital Equipment Co. Ltd., 1986.

Browne, Glenn J., Mitzi G. Pitts, and James C. Wetherbe. "Cognitive Stopping Rules for Terminating Information Search in Online Tasks." *MIS Quarterly*, vol. 31, no. 1, March 2007, pp. 89–104.

Bundorf, M. Kate, and Helena Szrek. "Choice Set Size and Decision Making: The Case of Medicare Part D Prescription Drug Plans." *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, vol. 30, no. 5, 2010, pp. 582–593.

Castleman, Benjamin L., and Lindsay C. Page. "Summer Nudging: Can Personalized Text Messages and Peer Mentor Outreach Increase College Going Among Low-Income High School Graduates?" *Journal of Economic Behavior & Organization*, vol. 115, 2015, pp. 144–160.

Chakraborty, G., R. Ettenson, and G. Gaeth. "How Consumers Choose Health Insurance." *Journal of Health Care Marketing,* vol. 14, no. 1, 1994, pp. 21–33.

Condliffe, Barbara F., Melody L. Boyd, and Stefanie DeLuca. "Stuck in School: How Social Context Shapes School Choice for Inner-City Students." *Teachers College Record,* vol. 117, no. 3, 2015, pp. 1–36.

Cronqvist, Henrik, and Richard H. Thaler. "Design Choices in Privatized Social-Security Systems: Learning from the Swedish Experience." *American Economic Review,* vol. 94, no. 2, 2004, pp. 424–428.

Deke, John. "Using the Linear Probability Model to Estimate Impacts on Binary Outcomes in RCTs," *OAH Evaluation Technical Assistance Brief #6,* December 2014.

Dynarski, Mark, Julian Betts, and Jill Feldman. "Applying to the DC Opportunity Scholarship Program: How Do Parents Rate Their Children's Current Schools at the Time of Application and What Do They Want in New Schools?" NCEE Evaluation Brief. Washington, DC: U.S. Department of Education, April 2016.

Gelman, Andrew. "Prior Distributions for Variance Parameters in Hierarchical Models." *Bayesian Analysis*, vol. 1, no. 3, 2006, pp. 515–533.

Gelman, Andrew, and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models Vol. 1*. New York, NY, USA: Cambridge University Press, 2007.

Gelman, Andrew. "Why We (Usually) Don't Have to Worry About Multiple Comparisons." *Journal of Research on Educational Effectiveness*, vol. 5, 2012, pp. 189–211.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis* (3rd ed.). Boca Raton, FL: Chapman and Hall/CRC Press, 2013.

Glazerman, Steven M., and Dallas Dotter. "Market Signals: Evidence on the Determinants and Consequences of School Choice from a Citywide Lottery." *Educational Evaluation and Policy Analysis*, April 2017. doi: 10.3102/0162373717702964

Glazerman, Steven M. "School Quality and Social Stratification: The Determinants and Consequences of Parental School Choice." Presented at the Annual Meeting of the American Educational Research Association, San Diego, CA, April 13–17, 1998.

Glazerman, Steven M. "Shopping for Schools: Mapping Choice Architecture in the Education Marketplace." Mathematica Working Paper Series, No. 59. Princeton, NJ: Mathematica Policy Research, December 2017.

Goodman, S. "Toward Evidence-Based Medical Statistics: 1: The P Value Fallacy." *Annals of Internal Medicine*, vol. 130, no. 12, June 1999a, pp. 995–1004.

Goodman, S. "Toward Evidence-Based Medical Statistics: 2: The Bayes Factor." *Annals of Internal Medicine*, vol. 130, no. 12, June 1999b, pp. 1005–1013.

Gross, Betheny, Michael DeArmond, and Patrick Denice. "Common Enrollment, Parents, and School Choice: Early Evidence from Denver and New Orleans." Bothell, WA: Center on Reinventing Public Education, May 2015.

Harris, Douglas, N., and Matthew Larsen. "What Schools Do Families Want (and Why)? School Demand and Information Before and After the New Orleans Post-Katrina School Reforms." New Orleans, LA: Education Research Alliance for New Orleans, January 2015.

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger. "Heterogeneous Preferences and the Efficacy of Public School Choice." NBER Working Paper No. 2145. Cambridge, MA: National Bureau of Economic Research, 2009.

Haxton, Clarissa L., and Ruth Curran Neild. "Parents' 'Hard' Knowledge of Admission Criteria and Acceptance in Philadelphia's High School Choice Process." *Journal of School Choice*, vol. 6, no. 4, 2012, pp. 483–503.

Hibbard, Judith H., Paul Slovic, Ellen Peters, and Melissa Finucane. "Strategies for Reporting Health Plan Performance Information to Consumers: Evidence from Controlled Studies." *Health Services Research*, vol. 37, no. 2, April 2002, pp. 291–313.

Hildon, Zoe, Dominique Allwood, and Nick Black. "Impact of Format and Content of Visual Display of Data on Comprehension Choice and Preference: A Systematic Review." *International Journal on Healthcare*, vol. 24, no. 1, 2012, pp. 55–64.

Hoekman Jr., Robert. *Designing the Moment: Web Interface Design Concepts in Action.* Berkeley, CA: Pearson Education, 2010.

Isakov, Leah, Andrew W. Lo, and Vahid Montazerhodjat. "Is the FDA Too Conservative or Too Aggressive?: A Bayesian Decision Analysis of Clinical Trial Design." NBER Working Paper No. 21499. Cambridge, MA: National Bureau of Economic Research, August 2015.

Jacobsen, Rebecca, Jeffrey Snyder, and Andrew Saultz. "Information or Shaping Public Opinion? The Influence of School Accountability Data Format on Public Perceptions of School Quality." *American Journal of Education*, vol. 121, no. 1, November 2014, pp. 1–27.

Jansen, S.J., A.M. Stiggelbout, M.A. Nooij, and J. Kievit. "The Effect of Individually Assessed Preference Weights on the Relationship Between Holistic Utilities and Nonpreference-Based Assessment." *Quality of Life Research*, vol. 9, no. 5, 2000, pp. 541–557.

Johnson, Eric J., Ran Hassin, Tom Baker, Allison T. Bajger, and Galen Treuer. "Can Consumers Make Affordable Care Affordable? The Value of Choice Architecture." *PLoS ONE*, vol. 8, no. 12, December 2013, p. e81521.

Jones, Gary, and Miles Richardson. "An Objective Examination of Consumer Perception of Nutrition Information Based on Healthiness Ratings and Eye Movements." *Public Health Nutrition*, vol. 10, no. 3, 2007, pp. 238-244.

Kahneman, Daniel. "Reference Points, Anchors, Norms, and Mixed Feelings." *Organizational Behavior and Human Decision Processes*, vol. 52, no. 2, March 1992, pp. 296–312.

Kahneman, Daniel, and Amos Tversky. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*, vol. 47, no. 2, March 1979, pp. 263–292.

Kruger, J., and David Dunning. "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessment." *Journal of Personality and*

Lewinsohn, S., and H. Mano. "Multi-Attribute Choice and Affect: The Influence of Naturally Occurring and Manipulated Moods on Choice Processes." *Journal of Behavioral Decision Making*, vol. 6, no. 1, 1993, pp. 33–51.

Loranger, Hoa, and Jakob Nielsen. *Prioritizing Web Usability*. Berkeley: New Riders Press, 2006.

March, James G. *Primer on Decision Making: How Decisions Happen*. New York, NY: Simon and Schuster, 1994.

McKelvie, S. J. "Graphic Rating Scales — How Many Categories?." *British Journal of Psychology*, vol. 69, 1978, pp. 185–202.

Nielsen, Jakob. *Coordinating User Interfaces for Consistency*. Boston: Academic Press, 1989.

Nielsen, Jakob. *Designing Web Usability: The Practice of Simplicity*. Indianapolis, IN: New Riders Publishing, 1999.

Nielsen, Jakob. *Usability Engineering*. San Francisco: Morgan Kaufmann, 1993.

Oppenheimer, Daniel M., Tom Meyvis, and Nicolas Davidenko. "Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power." *Journal of Experimental Social Psychology,* vol. 45, 2009, pp. 867–872.

Peters, E., N.F. Dieckmann, A. Dixon, J.H. Hibbard, and C.K. Mertz. "Less Is More in Presenting Quality Information to Consumers." *Medical Care Research and Review*, vol. 64, no. 2, April 2007, pp. 169–190.

Peters, Ellen, Nathan F. Dieckmann, Daniel Vastfjall, C. K. Mertz, Paul Slovic, and Judith H. Hibbard. "Bringing Meaning to Numbers: The Impact of Evaluative Categories on Decisions." Journal of Experimental Psychology: Applied, vol. 15, no. 3, 2009, pp. 213-227.

Russo, J.E. "The Value of Unit Price Information." *Journal of Marketing Research*, vol. 14, 1977, pp. 193–201.

Schneider, Mark, and Jack Buckley. "What Do Parents Want from Schools? Evidence from the Internet." *Educational Evaluation and Policy Analysis,* vol. 24, 2002, pp. 133–144.

Schneider, Mark, Paul Teske, Melissa Marshall, and Christine Roch. "Shopping for Schools: In the Land of the Blind, the One-Eyed Parent May be Enough." *American Journal of Political Science*, vol. 42, no. 3, 1998, pp. 769–793.

Schneider, Mark, Paul Teske, and Melissa Marschall. *Choosing Schools: Consumer Choice and the Quality of American Schools*. Princeton, NJ: Princeton University Press, 2002.

Simon, Herbert A. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics*, vol. 69, no. 1, February 1955, pp. 99–118.

Singh, Satyendra. "Impact of Color on Marketing." *Management Decision*, vol. 44, no. 6, 2006, pp. 783–789.

Srinivasan, V., and C.S. Park. "Surprising Robustness of the Self-Explicated Approach to Customer Preference Structure Measurement." *Journal of Marketing Research*, vol. 34, no. 2, 1997, pp. 286–291.

Stein, Mark L., and Sarah Negro. "The Readability and Complexity of District-Provided School-Choice Information." *Journal of Education for Students Placed at Risk*, vol. 20, no. 3, August 2015, pp. 199–217.

Stewart, Thomas, and Patrick J. Wolf. *The School Choice Journey: School Vouchers and the Empowerment of Urban Families*. New York, NY: Palgrave MacMillan, 2014.

Thaler, Richard. "Mental Accounting and Consumer Choice." *Marketing Science*, vol. 4, no. 3, 1985, pp. 199–214.

Thaler, Richard H., and Cass R. Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness.* New York: Penguin Group, 2009.

U.S. Department of Education. *What Works Clearinghouse Procedures and Standards Handbook, Version 3.0*. Washington, DC: U.S. Department of Education, 2014.

Valant, Jon. "Better Data, Better Decisions: Informing School Choosers to Improve Education Markets." Washington, DC: American Enterprise Institute, November 2014.

Wise, Steven L., and Xiaojing Kong. "Response Time Effort: A New Measure of Examinee Motivation in Computer-Based Tests." *Applied Measurement in Education*, vol. 18, no. 2, 2005, pp. 163–183.

U.S. Department of Education
Betsy DeVos
Secretary

Institute of Education Sciences
Mark Schneider
Director

October 2018

The report was prepared for the Institute of Education Sciences under Contract No. ED-IES-15-C-0048 with Mathematica Policy Research. The project officer is Meredith Bachman in the National Center for Education Evaluation and Regional Assistance.

Burnett, Alyson, Ira Nichols-Barrer, Steven Glazerman, and Jon Valant (2018). *Presenting School Choice Information to Parents: Technical Appendix* (NCEE 2019-4003). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

This report is available on the IES website at http://ies.ed.gov/ncee.

U.S. Department of Education
Institute of Education Sciences
550 12th Street, SW, Washington, DC 20202
(202) 245-6940
https://ies.ed.gov/ncee/