Assessing Fun Items’ Effectiveness in Increasing Learning of College Introductory Statistics Students: Results of a Randomized Experiment

Lawrence M. Lesser\textsuperscript{a}, Dennis K. Pearl\textsuperscript{b}, and John J. Weber, III\textsuperscript{c}

\textsuperscript{a}Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, Texas, USA; \textsuperscript{b}Department of Statistics, The Pennsylvania State University, University Park, Pennsylvania, USA; \textsuperscript{c}Department of Mathematics, Computer Sciences, and Engineering, Perimeter College at Georgia State University, Clarkston, Georgia, USA

\textbf{ABSTRACT}
There has been a recent emergence of scholarship on the use of fun in the college statistics classroom, with at least 20 modalities identified. While there have been randomized experiments that suggest that fun can enhance student achievement or attitudes in statistics, these studies have generally been limited to one particular fun modality or have not been limited to the discipline of statistics. To address the efficacy of fun items in teaching statistics, a student-randomized experiment was designed to assess how specific items of fun may cause changes in statistical anxiety and learning statistics content. This experiment was conducted at two institutions of higher education with different and diverse student populations. Findings include a significant increase in correct responses to questions among students who were assigned online content with a song insert compared with those assigned content alone.

\textbf{KEY WORDS}
CAUSEweb fun collection; Humor; Song; Statistics education research

1. Background

1.1. Literature on Fun

Fun (or at least, humor) has been studied in education for more than four decades, as captured in the review by Banas et al. (2011). A review of the literature on statistics fun was provided by Lesser and Pearl (2008) and Lesser et al. (2013), making connections to the Guidelines for Assessment and Instruction in Statistics Education (GAISE), American Statistical Association (2005), reviewing conceptualizations of fun, offering tips for classroom implementation, listing a score of delivery modalities (e.g., games, cartoons, songs, etc.) of fun, and listing sources of statistics fun (e.g., \url{https://www.causeweb.org/resources/fun/}). There are examples of the statistics education literature explicitly connecting specific pieces of fun to standard statistics content, such as the statistician hunters joke used in Peck, Gould, and Miller (2013, p. 69):

"The joke has meaning—and maybe is even funny—only if you understand a very important concept in statistics—bias. ... The hunting statisticians are happy because their shots are unbiased. One hunter shoots a meter above the duck (+1), the other a meter below (−1), and so, on average, they hit the duck."

Tews et al. (2015) “conceptualize fun instructor-initiated design and delivery elements as activities and interactions of an enjoyable, entertaining, humorous, or playful nature within a learning context” (p. 17). We believe a conceptualization should include the characteristics of surprise, humor, or play (SHoP) in the service of education. Lesser et al. (2013) reported on a survey of 249 instructors (and follow-up interviews with 16 of them) about their motivations and hesitations toward using fun modalities in the classroom. Because a consensus precise definition of fun has eluded researchers (see discussion in Lesser et al. 2013, sec. 1.1) and variations have also been noted in what statistics instructors count as fun (see Lesser et al. 2013, sec. 4.1.2), this article takes an operational definition of fun as simply the purposeful use of one of the modalities listed in Table 1 of Lesser and Pearl (2008) to accomplish lowering anxiety and/or increasing learning.

Experiments by Berk and Nanda (1998, 2006) and Garner (2006) indicate some evidence of effectiveness of fun items on increasing learning (as measured by test performance) and/or improving attitudes toward the lesson or instructor in statistics education in the modality of humor, but the statistics education literature has almost no studies involving other modalities. It is important not to assume that fun items using modalities other than humor will automatically work the same way as for humor for several reasons, including: (a) different modalities may occupy different places on the continuum of instructor’s perceived risk (in terms of embarrassment or lack of effectiveness) described by Berk (2003, 2005, 2006), showing a prepared cartoon is much lower risk than a live performance of a song, for example; (b) different modalities may have different interactions with students’ learning styles and/or cultural norms; (c) research on statistics instructors’ in-class usage of fun (Lesser et al. 2013) indicates some interaction between modality and gender: males were more likely than females to use humor while females were more likely than males to use games.
With respect to the modality of song, VanVoorhis (2002) conducted a quasieperiment on two sections of (psychology) statistics students and found that the section that sang jingle versions of three definitions did significantly better on related test items than the section that read the definitions in prose form. This result was limited in its generalizability by the lack of random assignment (although sections did have equal average GPAs) and the possibility of an instructor effect since it is unlikely that the instructor involved taught both sections otherwise identically (e.g., the reaction or direction of the ensuing conversation might differ in ways that set up additional content connections and that trajectory can have ripple effects on the tenor of that day’s lesson). A similar concern could even be expressed about the naturalistic experiment by Ziv (1988), in which students were randomly assigned to a 14-week introductory statistics course (taught by the same instructor) which had or did not have three or four jokes or cartoons inserted into each lecture.

There is also promising evidence of effectiveness of particular modalities in other disciplines. For example, song has been found by individual faculty to be a powerful vehicle for giving students possible benefits, such as improved recall, stress reduction, and motivation in a variety of disciplines, including: biology (Crowther 2006), psychology (Leck 2006), social studies (Levy and Byrd 2011), economics (McCough and Heinfeld 2012), food safety (McCurdy, Schmiege, and Winter 2008), mathematics (Lesser 2014), political science (Soper 2010), and sociology (Walczak and Reuter 1994). Likewise, usage of cartoons also spans many disciplines, including: social sciences (Ostrom 2004), English (Davis 1997; Brunk-Chavez 2004), geography (Kleeman 2006), computer science (Srikanw and Jakobsson 2007), geology (Rule and Auge 2005), physics (Para-Lez-Palacios and Vilchez-Gonzalez 2002), biology (Anderson and Fisher 2002), chemistry (Gonick and Criddle 2005), and mathematics (Greenwald and Nesterl 2004). Magic has also been used by college faculty in many disciplines, including: mathematics (Benjamin and Shermer 2006; Matthews 2008), physics (Featonby 2010), business and information systems (Elder, Deviney, and Mackinnon 2011), psychology (Solomon 1980; Bates 1991), organizational behavior (Krell and Dobson 1999), and probability/statistics (Lesser and Glickman 2009).

But because statistics is a context-rich, multifaceted discipline, it is safest not to assume that the results in other disciplines would automatically transfer to learning effects in statistics. For this reason, we would take advantage of the fact that for measuring attitudes or anxiety, there are statistics-specific instruments that we will describe shortly. Statistics anxiety is of particular importance here because it appears to be widespread among students and in society (Gal and Ginsburg 1994; Onwuegbuzie and Wilson 2003), has a higher likelihood than attitudes (which are considered more stable over time) to be changeable with appropriate interventions (Ziefel et al. 2008; Pearl et al. 2012), and is hypothesized to be affected by the use of fun items (Schacht and Stewart 1990; Berk and Nanda 1998; Chew and Dillon 2014). The literature we have discussed in this section motivated us to conduct a randomized experiment that simultaneously gauges the effect of fun on learning, attitudes, and anxiety while eliminating the instructor effect as a confounding factor.

1.2. Issues in Researching Fun

As discussed in the previous section, it is difficult to assess fun in a face-to-face classroom under controlled, replicable conditions. One way to address this is to evaluate the efficacy of fun in an online course component (e.g., using song vs. not using song or examining other specific modalities in an online environment). In this context, a simple experimental design would have a controlled fixed product, such as a video or an electronic reading that can be electronically duplicated, and then a fun item can be inserted into only one of the two versions. These versions can, therefore, truly be otherwise identical. This design was implemented by Garner (2006), who randomly assigned $n = 117$ undergraduates to review three 40-minute lecture videos on statistics research methods, with or without content-relevant jokes inserted. This setup ensured that all other aspects of the lectures (in distance education format) were controlled to be identical. Students in the group with the humor condition gave significantly higher ratings (each $p$-value was $< 0.001$) in their opinion of the lesson, how well the lesson communicated information, and quality of the instructor. Even more importantly, the group with the humor condition also recalled and retained more information on the topic (this $p$-value was also $< 0.001$). Özdogr and McMorris (2013) had $n = 156$ undergraduates study six short readings, where three had content-related cartoons inserted. The students appeared to have a more favorable attitude toward the cartoons themselves, believing them to be helpful to their learning, but the data did not show a significant direct effect on student performance.

With the advent of customizable open-source textbooks (e.g., http://openintro.org/stat/, http://pearsonhigheredolutions.com/collections, and http://openstax.org/details/introductory-statistics), it has only become easier to conduct a study where a treatment group receives an identical copy of a book with one or more fun artifacts inserted. Removing the variable of instructor effect has the secondary benefit of addressing the second most common hesitation on using fun mentioned by college instructors (see sec. 4.3.1 of Lesser et al. 2013), namely, lacking the necessary skill or talent. If inserted fun items can be shown to be effective even when they are simply inserted into the clinical environment of a learning management system (LMS), a video, or an e-book, then any instructor could provide these materials for his/her students, and it is arguably the case that if the instructor used these items during live in-class instruction (with the advantage of all of the emotional supports and cues of in-person interaction as well as the dynamics of a room full of people who might feed off of each other’s “fun instincts”), the success would be no less. And while not every instructor may be ready to try the high-risk action of a live performance of an original song, there is little risk or talent needed to browse the free searchable collection at https://causeweb.org/resources/fun/ for a song on the statistics topic of interest and just press PLAY to share the recording in class.

Having online readings in this type of study also removes the possible obstacles of needing in-class time (or preparation) and offers a collection of fun-enhanced minireadings that could be readily used in a standardized fashion by any instructor teaching an online (or hybrid) statistics course, a type of course which is becoming increasingly common in higher education.
In particular, Blair, Kirkman, and Maxwell (2010) reported that 39% of statistics departments at four-year institutions and 88% of two-year colleges used distance learning courses and these percentages are almost certainly even higher now. Hence, it is important that, beyond the potential for use in face-to-face courses, the medium of fun items in the experiment discussed in this article also applies to hybrid and online courses. LoSchiavo and Shatz (2005) conducted what they claim is the first study of humor incorporated into an online course (on general psychology, using the Blackboard Learning System). After randomly assigning students to either the humor-added or regular sections of the online course, they found that “humor significantly influenced student interest and participation, but had no effect on overall course performance.” (Shatz and LoSchiavo 2006). This study did not focus on a statistics course, unlike the design reported by us now. Our study is specifically designed to ameliorate the issues Shatz and LoSchiavo (2006) discuss regarding some of the ways in which fun (or, at least, humor) is constrained by a classroom setting, and constrained further still in an online setting, where “humor cannot be embellished by nonverbal cues or easily retracted.”

### 1.3. Hypotheses

Our main research hypothesis is that students exposed to statistics fun items will experience improved learning in statistics and that there may be a difference in the effect on learning between modalities. A secondary research hypothesis was that students exposed to statistics fun items will experience lower statistics anxiety as measured by the Statistics Anxiety Measure (SAM; Earp 2007). We also evaluated attitudes toward statistics, as measured by the Survey of Attitudes Toward Statistics (SATS; Schau et al. 1995). The SATS is the most used student attitudes instrument in the statistics education research literature (see http://www.evaluationandinstructorstats.com/references.html for a list of many studies on or that have used the SATS). However, due to the stability of SATS, except under extreme interventions, we expected a negligible potential for pre-post change in these resistant student attitudes toward statistics (attitudes generally being considered to be a “trait” rather than a “state”; Pearl et al. 2012). Thus, our a priori research hypotheses were that student learning would improve and anxiety would be reduced in contrast to student attitudes, which are more stable.

### 2. Methodology

#### 2.1. Settings

In order to ascertain whether any effects of fun items can cut across diverse populations, two institutions were chosen with different and diverse student populations. As Table 1 indicates, there is also variation between the institutions in other ways, such as the students’ majors and course textbooks.

#### 2.2. Identification of Items

The largest searchable collection of fun items in statistics, spanning many fun modalities, is at https://www.causeweb.org/resources/fun/, housed within the digital library of the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE). This digital library for college-level statistics instructors is part of the National Science Foundation’s (NSF) National Science Digital Library (NSDL) system and is also affiliated with Multimedia Education Resource for Learning and Online Teaching (MERLOT) and the Mathematical Sciences Digital Library (MathDL). One of CAUSEWeb’s most-visited resources, the fun collection contains over 500 items.

| Topic of minireading | Sullivan (2014) correlate | Utts (2005) correlate | Inserted fun item id# (modality) |
|----------------------|---------------------------|----------------------|--------------------------------|
| Categorical versus quantitative variables | 1.1 (p. 9) | 3.5 (pp. 46–47) and 4.2 (p. 61) | 486 (poem) |
| Margin of error | 1.1 (p. 6) and 9.1 (p. 392) | 4.3 (p. 62) | 344 (song) |
| Interval versus point estimate | 9.1 (p. 391) | 4.3 (p. 62) | 419, 216 (cartoons) |
| Simple random sampling | 4.4 (p. 64) | 3 (cartoon) |
| Random versus haphazard | 1.3 (p. 22) | 4.4 (p. 64), etc. | 468 (cartoon) |
| Inadequacy of mean without variation | 3.2 (p. 131) | 7.1 (p. 125) | 181 (quote) |
| Mean versus median | 3.1 (pp. 120–1) | 7.5 (p. 137) | 258 (cartoon) |
| Standard score | 3.4 (p. 154) | 8.3 (p. 151) | 85 (song) |
| Correlation and slope | 4.1 (p. 181) | 10.4 (p. 189) | 487 (song) |
| Correlation versus causation | 4.1 (p. 185) | 11.3 (p. 207) | 5 (cartoon) |
| Equiprobability bias | 5.1 (p. 239) | 16.2 (p. 299) | 506 (song) |
| Multiplicity | Not used | 16.5 (p. 305) | 49 (song) |
| Sample size versus margin of error | 9.1 (p. 399) | Not used | 350 (joke) |
| p-value | 10.2 (p. 438) | Not used | 86 (song) |

**Table 2.** Fun items used in the student randomized experiment.

**NOTE:** The CAUSEweb id# can be used to access the fun item by inserting it after the equals sign in https://www.causeweb.org/resources/fun/db.php?id=
The authors identified a set of 14 fun items (see Table 2) from the CAUSEweb collection that correspond to a representative and diverse set of learning objectives in the introductory statistics courses of the institutions participating in the study. A condition for inclusion in this set was that the fun items not merely be related to the topic, but could actually stimulate thinking about a learning objective in the introductory course. The latter criterion ruled out, for example, a joke that simply made a pun on a statistics word but was not instructive on its own in helping a student understand or remember statistics content.

The research team examined the entire CAUSEweb fun collection, searching for items that aligned well with the operational definition of fun and with the learning objectives in the courses at the institutions participating in the study. Most of the items selected for the study were songs and cartoons, in part because they are the second and third most abundant categories in the CAUSEweb collection. Quotations, the most abundant category in the collection, were deemphasized in the selection process and only one was chosen based on its use of humor and its connection to the learning objective at hand. In general, to keep the number of minireadings manageable for the students, the number of fun items had to be kept manageable as well. Thus, it was felt that having the distribution of categories include two categories tested more extensively would allow for modality-specific effects to be judged more powerfully than having, say, six or seven categories with only two items each. The Banas et al. (2011) review paper classifies types of humor as either appropriate for the classroom, inappropriate, or dependent on context. For example, denigrating others or basing humor on a person’s demographic characteristics are viewed as inappropriate for the classroom, while doing impersonations or using (occasional) sarcasm would depend on context. Thus, we deliberately chose our fun items to avoid this pitfall identified in the literature.

2.3. Example of Insert

Figure 1 shows the item for teaching that correlation and slope (of the regression line for the same scatterplot) share the same sign. The “control version” ends at the solid line before the song.

2.4. Student-Randomized Experiment

In fall 2013, quantitative data were collected from 53 students who completed the course and gave consent at the two-year college and 115 students at the medium-sized university. In the first two weeks of the semester, all the students were asked to take the pretest versions of the SATS-36 (Schau et al. 1995) and the SAM (Earp 2007). In the last two weeks of the semester and prior to the final exam, all the students were asked to complete the posttest versions of SATS-36 and SAM. Half of the students from every section were randomized to have fun items inserted into content minireadings, like the one in Figure 1 in sec. 2.3, assigned by the instructors. The students accessed the readings via the LMS in their course. The minireadings were generally one to two paragraphs long and self-contained, so that they worked whether or not a fun item was inserted. To give the students incentive and encouragement to access the readings, the instructors informed the students that each minireading would have an associated multiple-choice question on a midterm or final. For example, the embedded question associated with the item in Figure 1 appears in Figure 2.

2.5. Human Subjects Considerations

All students participating in this experiment were 18 years or older and gave informed consent through procedures approved by each institution’s Institutional Review Board. For the SAM and SATS surveys, confidentiality was maintained through a secure server operated by CAUSE at The Ohio State University.
without receiving any personal student information from the other two educational institutions. In addition, all student names were present on that server only in an encrypted format and could not be retrieved by anyone at Ohio State. At the conclusion of data collection (and after grades were completed) at the two institutions participating in the student-randomized study, the SAM and SATS data were sent from the server to the two institutions, where they were merged with locally-collected data (including GPA, age, gender, and race/ethnicity).

3. Results of the Experiment

Data were analyzed to see if introductory statistics students exposed to fun modalities, such as CAUSEweb.org cartoons or songs inserted into 14 otherwise conventional online short content items (12 of which were used at one institution and 13 at the other, as listed in Table 2) would perform better on related embedded multiple-choice exam questions, display greater improvement in attitudes toward statistics (measured by the SATS-36; Schau et al. 1995), or greater decrease in statistics anxiety (measured by the SAM; Earp 2007) over a semester.

3.1. Results on Learning

Six of the embedded test items involved the use of songs to teach introductory material to a randomly selected group of students (four items used at both schools and one unique item at each school). As shown in Table 3a, students exposed to the song inserts in their LMS performed significantly better on related multiple-choice items embedded on midterm and final exams. In fact, all six of these song items showed a higher percent of correct answers among students who viewed the lesson embedded with the song, compared with the control students who saw the lesson alone. Overall, students randomized to the fun group corresponding to readings accompanied by songs, answered those embedded questions correctly an average of 50.0% of the time. In comparison, students randomized to the lessons without songs got them correct an average of 42.3% of the time (two-tailed p > 0.04). The use of cartoons showed a somewhat negative though not significant effect (p > 0.22) while the other items (a joke used at only one institution, a poem, and a quote) showed a somewhat positive but not significant difference between experimental and control groups on test item performance (p > 0.32) (see Table 3b).

![Figure 2. Assessment item corresponding to minireading in Figure 1.](image)

### Table 3a. Percent correct on statistical concept assessment, by with/without (song) inserts.

| Topic of reading/insert (CAUSEweb id#) | Without song insert n = 88 | With song insert n = 80 | Difference |
|--------------------------------------|---------------------------|------------------------|------------|
| Margin of error topic (334): ME decreases with n | 57.3% | 61.3% | 4.0% |
| Margin of error topic (334): ME decreases at rate \(\sqrt{n}\) | 9.1% | 10.0% | 0.9% |
| Standard score (85) | 62.5% | 73.8% | 11.3% |
| Correlation and slope (487) | 60.2% | 73.8% | 13.6% |
| Equiprobability bias (506) | 40.9% | 50.0% | 9.1% |
| Multiplicity (49) | 36.1% | 37.0% | 0.9% |
| p-value (86) | 44.4% | 50.0% | 5.6% |
| Overall song | 42.3% | 50.0% | 7.7% |

90% CI on total difference (1.4%, 14.0%); p-value ≈ 0.04

NOTES: 1Results only from university with n = 61 without insert and n = 54 with insert.
2Results only from two-year college with n = 27 without insert and n = 26 with insert.

### Table 3b. Percent correct on statistical concept assessment, by with/without (nonsong) inserts.

| Topic of reading/insert (CAUSEweb id#) | Modality of insert | Without insert n = 88 | With insert n = 80 | Difference |
|--------------------------------------|-------------------|-----------------------|------------------|------------|
| Interval versus point estimate (419, 216) | Cartoon | 73.9% | 63.8% | −10.1% |
| Simple random sampling (3) | Cartoon | 61.4% | 48.8% | −12.6% |
| Random versus haphazard (468) | Cartoon | 59.1% | 52.5% | −6.6% |
| Mean versus median (258) | Cartoon | 61.4% | 55.0% | −6.4% |
| Correlation versus causation (5) | Cartoon | 42.0% | 52.5% | 10.5% |
| Overall cartoon | Cartoon | 59.5% | 54.5% | −5.0% |
| Categorical versus quantitative variables (486) | Poem | 17.0% | 20.0% | 3.0% |
| Inadequacy of mean without variation (181) | Quote | 22.7% | 27.5% | 4.8% |
| Sample size versus margin of error (350) | Joke | 40.9% | 50.0% | 9.1% |
| Overall other | Joke | 22.6% | 27.4% | 4.8% |

NOTE: *Results only from two-year college with n = 27 without insert and n = 26 with insert.
The two-sided $p$-value and 90% confidence interval for the effect of songs were computed using a $2 \times 2 \times 2 \times 3$ ANOVA with the dependent variable of “proportion of embedded questions for song items answered correctly” and with the independent variables of “whether or not a student was randomly assigned to be exposed to a fun item,” age (18–24 or 25+), institution, gender, and GPA (first semester student with no GPA; above the institutional median; or below the institutional median). The choice of 24 for the upper bound of the lower age category was chosen to follow the commonly used benchmark for “traditional” versus “nontraditional” students (see, e.g., https://nces.ed.gov/pubs/web/97578e.asp). The latter four variables did not show significance with the dependent variable. Race/ethnicity was also not a significant factor, as it is essentially a surrogate for the institutional site. The use of other modalities in this study did not show significance. Although the response variable was not on a continuous scale, a normal probability plot of the residuals showed a very linear pattern.

3.2. Results on Attitudes and Anxiety

The number of students completing both the pretest and posttest inventories did not provide adequate power to detect post-test differences on SATS and SAM because the instructors asked, but could not require, the students to complete the inventories. Both SATS and SAM had a lower response rate for the posttest version of the inventory as some students withdrew from the course or perhaps became less likely to participate in this optional activity near the end of the term. Specifically, for the SATS, 117 students took the pretest, 88 took the posttest, and only 54 took both. For the SAM, 121 took the pretest, 112 took the posttest, and only 68 took both. Since a small proportion of students completed both pre- and postversions of the inventories, the authors analyzed posttest data only. In particular, we note that the $R^2$ values between pre and posttest scores (i.e., the percent reduction in variance achievable by adjusting for pretest levels) were lower than the percent decrease in variance that was achieved because of the higher posttest sample sizes for all of the SAM subscales and for four of the six SATS subscales. The remaining two SATS subscales would have provided less than an 8% improvement by adjusting for the pretest values.

As expected (see sec. 1.3), each of the six subscales of the Student Attitudes Toward Statistics instrument changed little over the period of the study among students in either the treatment or control groups (Table 4). Student anxiety, as measured by the SAM instrument, also did not show a significant effect of fun, though the results had a trend toward a beneficial effect in this small sample (Table 5).

4. Discussion

This work was undertaken to examine whether brief fun items would improve student learning and reduce anxiety in contrast to student attitudes, which are known to be more stable. We found improvements in student learning arising from songs (though not from other modalities) and a weak improvement in student anxiety.

As mentioned at the end of sec. 2.2, after observing that learning gains seen with songs were not replicated for other modalities, we conjectured that this result may be due to song items being (a) more memorable or (b) more engaging/active. We now discuss each of these two conjectured factors.

Song may be the modality that has the strongest mnemonic potential (see sec. 4.3) and may also be inherently more able to involve sustained engagement. For example, while a cartoon with its one-liner caption can be passively browsed in just seconds, a song may take a few minutes to play from the sound file and meanwhile lyrics remain in view to be read or even sung along with (as the minireading explicitly invited viewers to do). This relates to “time on task,” a variable that is “positively related to persistent and subsequent success in college,” especially for online education (National Survey of Student Engagement [NSSE] n.d.; Chickering and Ehrmann 1996; Pri-neas and Cini 2011). Another rationale that favors components of songs with their visual (lyrics) and auditory (sound file) components is that students can take in an increased amount of information through multiple input channels with independent capacities (Petty 2010). Yet another consideration is that listening to music is associated with the pleasurable release of dopamine in the brain (Salimpour et al. 2011).

It is possible that fun items may have different levels of “activeness,” a factor widely recognized in the literature. It is the third principle of Chickering and Gamson’s (1987) Seven Principles for Good Practice in Undergraduate Education and the fourth recommendation of the six GAISE recommendations (ASA 2005). The GAISE recommendation states that active learning “allows students to discover, construct, and understand important statistical ideas and to model statistical thinking” and “often engage students in learning and make the learning process fun” (ibid, p. 18). There is also broad empirical support for active learning in the STEM education research literature, reflected in the metaanalysis (Freeman et al. 2014) of 225 studies that concluded that active learning increases performance on exams. In the current study, a song that asked students to sing along with a recording might be considered highly active while a cartoon where students were not asked to do more than view the image might be considered less active.

### Table 4. Postcourse means (± standard error) for SATS subscales (higher numbers are better).

| Measure/subscale | Control, $n = 44$ | Fun, $n = 44$ | Difference |
|------------------|------------------|--------------|------------|
| Affect           | 4.11 ± 0.19      | 4.18 ± 0.21  | 0.07 ± 0.28|
| Cognitive        | 4.66 ± 0.16      | 4.73 ± 0.22  | 0.06 ± 0.27|
| Value            | 4.48 ± 0.15      | 4.46 ± 0.22  | −0.02 ± 0.26|
| Difficulty       | 3.74 ± 0.15      | 3.75 ± 0.17  | 0.01 ± 0.23|
| Interest         | 3.69 ± 0.21      | 3.89 ± 0.28  | 0.20 ± 0.35|
| Effort           | 6.02 ± 0.20      | 6.06 ± 0.17  | 0.04 ± 0.26|

### Table 5. Postcourse means (± standard error) for SAM and its subscales (lower numbers are better).

| Measure/subscale | Control, $n = 59$ | Fun, $n = 53$ | Difference |
|------------------|------------------|--------------|------------|
| Anxiety          | 8.32 ± 0.44      | 7.53 ± 0.45  | −0.79 ± 0.63|
| Attitude toward class | 7.17 ± 0.26 | 6.40 ± 0.28  | −0.77 ± 0.39|
| Fearful behavior  | 12.86 ± 0.42     | 12.64 ± 0.46 | −0.22 ± 0.62|
| Attitude toward math | 12.46 ± 0.73  | 12.25 ± 0.64 | −0.21 ± 0.97|
| Performance      | 9.80 ± 0.36      | 10.26 ± 0.39 | 0.46 ± 0.53|
| SAM total        | 50.61 ± 1.51     | 49.08 ± 1.71 | −1.53 ± 2.28|
With respect to the student attitudes measured by the SATS instrument, our results align with the common finding that such attitudes are difficult to modify with a single class related intervention (Ziegfler et al. 2008; Pearl et al. 2012). Statistical anxiety, as measured by SAM and other instruments, is more of an alterable state and the use of humor, in particular, is cited by Chew and Dillon (2014) as one of the five most important interventions that can be used to alleviate its effects. This is especially true of the components of anxiety dealing specifically with anxiety and related attitudes about the course the student is enrolled in, as opposed to the nonclass-related anxieties and attitudes about the discipline of statistics in everyday life. Interestingly, these were the subcomponents of SAM that showed the most change in our experiment—though the changes were not large and even the change in the overall instrument from the control to the treated group were not statistically significant. While the 3% (= 1.53/50.61) change seen in the overall SAM values might not be of any practical significance, the changes of around 10% seen in the subscales for anxiety and attitudes toward the class are likely to be of practical importance. Hence, replicating the experiment with an adequate power to detect this effect is needed. Note that all items included in the study can be considered humor/levity/playfulness in order to lift up the spirits of students and are, thus, aligned with the intervention suggested by Chew and Dillon (2014). Readers may find further contextualization of how these populations relate to fun in the qualitative analysis reported by Lesser et al. (2014) and Lesser and Reyes (2015).

In the student-randomized experiment, why did songs do well and appear to do better than, say, cartoons? One possible factor is that it takes longer to hear a song than to glance at a cartoon. Governor, Hall, and Jackson (2013) discussed how songs prolong listeners’ engagement with concepts and activate multiple neural connections and pathways. A related factor is that song is simply a more active modality by engaging both visual and auditory senses and the experiment’s minireading invited the student to “play it again and sing along.” Finally, song is arguably the modality that has the strongest mnemonic potential due to all the dimensions (imagery, rhyme, rhythm, melody, emotion, etc.) it has to cue or chunk information (Calvert and Tart 1993; Wallace 1994).

4.1. Limitations

While this experiment has yielded a few promising and interesting findings, these should be viewed in context with the study’s limitations. The most serious limitation is that we did not get enough students in the student-randomized experiment to participate completely. Since the instructors at each institution could not require students to complete the SAM and SATS inventories, the result was an insufficient sample to compare prescores to postscores.

A further limitation is that we were not able to measure engagement with the fun items in the LMS. Even if we had been able to readily record, for example, how much time students spent on a particular page within the LMS, this is an imperfect proxy for engagement with the item (e.g., we do not know how much of that time their attention was truly focused on the screen). In the case of lessons with an embedded song, there was an additional step of clicking on a hyperlink to the sound file, which was not recorded in the log file of the LMS.

The lessons were presented to the students throughout the semester. However, at the two-year college, five of the lessons covered five consecutive sections (typically covered over 1–2 class periods) of the curriculum during the first half of the semester. Another four of the lessons covered four consecutive sections (typically covered over 1–2 class periods) of the curriculum during the second half of the semester. It is unknown whether the number/timing of the minilessons required over such a short time period within the semester may have had an effect on the students’ engagement with the lessons. It is worth noting (as a limitation as well as a direction for future research) that we also do not know how the minilessons’ online format (as opposed to a printed textbook format, as used by Özdoğru and McMorris 2013) may have impacted the results.

Also, due to IRB protocol restrictions and variations in how the two institutions recorded information, we could only crudely control for students’ GPA, major, and ethnicity. For example, the university had no GPA data for 34% of students who were in their first semester (freshman entry or transfer), while the students’ GPA at the two-year college was collected after the semester ended and thus included the grades of the semester during which the study was conducted.

The experimental design did not inundate the curriculum with fun—so, the minimal number of items would have little chance to affect attitudes, which are only slightly modifiable. Anxiety is modifiable (Chew and Dillon 2014)—and may also be related to the degree of activeness of students (i.e., the fun items will have little chance of improving learning, and possibly even attitudes, if they are not presented in a pedagogically sound, active manner).

Another limitation (noted in a conversation between one of the authors and an attendee at the authors’ poster presented at ICOTS9) may be the placement of the fun item within the lesson. In each of the lessons with an embedded fun item, the fun item was placed either at the end of the lesson or immediately prior to a prompt to sing along or think aboutthe underlying statistical concept contained in the fun item. A fun item placed at the beginning of the lesson may have the biggest potential to reduce student anxiety and therefore, most help the student complete the remainder of the lesson.

4.2. Directions for Future Research

One direction for future research would be to repeat the study with a larger sample size. A larger sample size may be acquired by encouraging instructors to emphasize the importance of the lessons more frequently in class. Additionally, the study can be repeated using additional colleges and universities. Finally, it is likely that online classes, where students would complete all experimental assessments as part of their normal online work, may result in a higher participation rate in a future study.

Since attitudes are more difficult to change (Pearl et al. 2012), future studies should focus on student anxiety using attitudes as a precourse covariate. There is also a possibility to get more insight on anxiety by not relying solely on self-reports, but also including “some real-time physiological measures indexing stress and anxiety” (Wang et al. 2014, p. 7). An
example might be the use of actigraphy software on smart watches that tracks blood pressure and heart rate.

Future studies should vary the placement of the fun item in the lesson. The fun items should be embedded sometimes at the beginning, sometimes at the end, and sometimes in the middle. This may be helpful to determine an optimal location for the fun item to maximize the effect of the fun item on student learning and student anxiety reduction.

Since half of the fun items used within this research happened to be songs (see discussion in sec. 2.2), future research needs to assess other modalities of fun identified by Lesser and Pearl (2008). For example, magic might be particularly useful in learning concepts of probability as an underpinning for significance testing (Lesser and Glickman 2009). Students could experience a magic trick (either in a face-to-face classroom, in an online video from the CAUSEweb collection, or in an online applet) and be prompted to calculate the probability of the correct prediction (under the implicit hypothesis that the magic trick involved independence and fairness with its dice, cards, etc.). Since the CAUSEweb collection of fun resources has more than 500 fun items, future research might include additional items from the collection to complement the result from this research. Future research should have follow-up interviews to focus specifically on why students found some items engaging and helpful and other items not. Other aspects of each item might be evaluated and examined explicitly as explanatory variables, such as the use of rhyme or students’ perception of “memorableness” or “funniness,” or the degree to which an item encourages more interaction by the student.

Acknowledgments

This work was presented in part at the 2014 ICOTS and eCOTS conferences. We also thank those who gave advice or assistance with the data collection process: Tracy Adkins, Francis Biney, Cynthia Box, Donna Brouillette, Guillermo Castillo, Gerard Cochrane, Walter Copenhaver, Larry Coty, Ron Graham, William Griffin, Didier Hernandez, Luis Hernandez, Yin Lin, Nigel Scott, Justin Slauson, and Jianchun Zhang. We also greatly appreciate the very helpful editor and referee feedback provided by JSE.

Funding

This work was supported in part by Project UPLIFT: Universal Portability of Learning Increased by Fun Teaching NSF/EHR/DUE-1141261 (PSU), 1141261 (OSU), 595 1140690 (UTEP), 1140592 (GPC). The authors are the grant’s PIs.

References

American Statistical Association (2005), Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report, Alexandria, VA: ASA. Available at http://www.amstat.org/education/gaise/GaiseCollege_Full.pdf.

Anderson, D. L., and Fisher, K. M. (2002), “Biological Concept Cartoons: A Constructivist Tool for Dealing with Alternative Conceptions,” Paper presented at the annual meeting of the National Science Teachers Association, San Diego, CA.

Banas, J. A., Dunbar, N., Rodriguez, D., and Liu, S.-J. (2011), “A Review of Humor in Educational Settings: Four Decades of Research,” Communication Education, 60, 115–144.

Bates, J. A. (1991), “Teaching Hypothesis Testing by Debunking a Demonstration of Telepathy,” Teaching of Psychology, 18, 94–97.

Benjamin, A. and Shermer, M. (2006), Secrets of Mental Math: The Mathemagician’s Guide to Lightning Calculation and Amazing Math Tricks, New York: Random House.

Berk, R. A. (2003), Professors are From Mars, Students are From Snickers: How to Write and Deliver Humor in the Classroom and in Professional Presentations, Sterling, VA: Stylus. (2005–2006). “Laughterpiece Theatre: Humor as a Systematic Teaching Tool,” Essays on Teaching Excellence: Toward the Best in the Academy, 17, 1–8. Available at http://podnetwork.org/content/uploads/V17-N2-Berk.pdf.

Berk, R. A., and Nanda, J. P. (1998), “Effects of Jocular Instructional Methods on Attitudes, Anxiety, and Achievement in Statistics Courses,” HUMOR: International Journal of Humor Research, 11, 383–409.

Berk, R. A., and Nanda, J. P. (2006), “A Randomized Trial of Humor Effects on Test Anxiety and Test Performance,” HUMOR: International Journal of Humor Research, 19, 425–454.

Blair, R., Kirkman, E. E., and Maxwell, J. W. (2010), Statistical Abstract of Undergraduate Programs in the Mathematical Sciences in the United States: Fall 2010 CBMS Survey, Washington, DC: American Mathematical Society. Available at http://www.ams.org/profession/data/cbms-survey/cbms2010-Report.pdf.

Bruen-Chavez, B. L. (2004), “What’s So Funny about Stephen Toulin?: Using Cartoons to Teach the Toulinan Analysis,” Teaching English in the Two-Year College, 32, 179–185.

Calvert, S. L., and Tart, M. (1993), “Song Versus Verbal Forms for Very-Long-Term, Long-Term, and Short-Term Verbatim Recall,” Journal of Applied Developmental Psychology, 14, 245–260.

Chew, P. K. H., and Dillon, D. B. (2014), “Statistics Anxiety Update: Refining the Construct and Recommendations for a New Research Agenda,” Perspectives on Psychological Science, 9, 196–208.

Chickering, A. W., and Ehrmann, S. C. (1996), “Implementing the Seven Principles: Technology as a Lever,” American Association of Higher Education Bulletin, 49, 3–6. Available at http://www.aaehe.org/aahe/articles/sevenprinciples.htm.

Chickering, A. W., and Gamson, Z. F. (1987), “Seven Principles for Good Practice in Undergraduate Education,” AAHE [American Association for Higher Education] Bulletin, 39, 3–7. Available at http://www.aaehe.org/aahe/articles/sevenprinciples1987.htm.

Crowther, G. (2006), “Learning to the Beat of a Different Drum,” Connect, 19, 11–13.

Davis, R. S. (1997), “Comics: A Multidimensional Teaching Aid in Integrated-skills Classes,” Studies in Social Sciences and Humanities, Available at http://www.esi-lab.com/research/comics.htm.

Earp, M. A. (2007), “Development and Validation of the Statistics Anxiety Measure,” unpublished dissertation, University of Denver. Available at http://aase-web.org/documents/dissertations/07.Earp.Dissertation.pdf.

Elder, K. L., Devine, D. E., and MacKinnon, R. J. (2011), “A Method to Increase Student Comprehension with Illusions,” in 2011 Proceedings of the Society of Business, Industry, and Economics Annual Meetings (Ed) Vivek Bhargava, pp. 78–84.

Featonby, D. (2010), “Magic Physics?: Physics Education, 45, 24–31.

Freeman, S. Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., and Wenderoth, M. P. (2014), “Active Learning Increases Student Performance in Science, Engineering, and Mathematics,” Proceedings of the National Academy of Sciences of the United States of America, 111(23), 8410–8415. Available at http://www.pnas.org/content/111/23/8410.full.

Gal, I., and Ginsburg, L. (1994), “The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework,” Journal of Statistics Education, 2. Available at http://www.amstat.org/publications/jse/v2n2/gal.html.

Garner, R. L. (2006), “Humor in Pedagogy: How Ha-Ha Can Lead to Aha!,” College Teaching, 54, 177–180.

Gonick, L., and Criddle, C. (2005), The Cartoon Guide to Chemistry, New York: Harper Collins.

Governor, D., Hall, J., and Jackson, D. (2013), “Teaching and Learning Science through Song: Exploring the Experiences of Students and Teachers,” International Journal of Science Education, 35, 3117–3140.

Greenwald, S. J., and Nestler, A. (2004), “r dr r: Engaging Students with Significant Mathematics from The Simpsons,” PRIMUS, 14, 29–39.
