HOW WE TEACH | *Generalizable Education Research*

Is memorization the name of the game? Undergraduates' perceptions of the usefulness of physiology songs

Gregory J. Crowther,[1,2] Jason Wessels,[3,4] Lawrence M. Lesser,[5] and Jennifer L. Breckler[6]

[1]*Department of Life Sciences, Everett Community College, Everett, Washington;* [2]*Division of Biological Sciences, University of Washington Bothell, Bothell, Washington;* [3]*Science Department, Eastlake High School, Sammamish, Washington;* [4]*College of Education, University of Washington Bothell, Bothell, Washington;* [5]*Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, Texas; and* [6]*Department of Biology, San Francisco State University, San Francisco, California*

Crowther GJ, Wessels J, Lesser LM, Breckler JL. Is memorization the name of the game? Undergraduates' perceptions of the usefulness of physiology songs. *Adv Physiol Educ* 44: 104–112, 2020; doi:10.1152/advan.00112.2019.—The possible benefits of using music to enhance learning of STEM content are numerous, diverse, and largely unproven. We sought to determine which (if any) of these possible benefits are commonly experienced by undergraduate students and are thus especially worthy of further investigation. Four hundred ninety-three students in nine physiology courses at two midsized American universities rated the usefulness of short instructor-penned mathematical physiology songs and explained in their own words why each song would or would not be a useful study aid. The students collectively perceived the usefulness of each song to depend on both academic factors (e.g., the lyrics' clarity or relevance to the course) and aesthetic values (e.g., the appeal of the rhythm or the quality of the singing). Most strikingly, although students' free responses were brief (median length: 18 words in study *phase 1*, 16 words in study *phase 2*), nearly one-half of them (1,039 of 2,191) concerned memory, suggesting that many students see educational songs primarily as mnemonic devices. A second major theme of students' comments concerned the conciseness and information density of the songs. Though all 10 songs were brief, lasting 17–54 s, students seemed to prefer shorter songs (perhaps better called "jingles"). This first-of-its-kind data set on student perceptions of educational songs should inform the creation and usage of such songs, as well as further research on their possible value.

edutainment; music videos; STEAM; student perception

INTRODUCTION

A recent National Academies report documents emerging pedagogical interest in integrating STEM (science, technology, engineering, and mathematics) subjects with the arts (30), a movement often referred to as STEAM (23). Among the many possible ways of adding art to STEM, or vice versa, a relatively popular option is using the musical arts to teach STEM content, an approach that has its own annual online conference (18).

In theory, music could enhance STEM learning through many different mechanisms, depending on the instructor, students, course, and other factors. These mechanisms, summa-rized in Table 1, are compatible with general models of teaching and learning, such as the "5E" (engage-explore-explain-elaborate-evaluate) model (35). For example, in the "explore" phase, students might note "mysterious" aspects of a content-rich song that do not yet make sense to them, while in the "evaluate" phase, students might demonstrate understanding by editing or critiquing an existing song (1–3, 7, 9–11, 13, 17, 26, 34, 36, 37).

While some studies have sought to demonstrate positive effects of music on learning (5, 13, 20, 25, 33), hardly any have formally tested individual mechanisms (e.g., those in Table 1) as falsifiable hypotheses. We propose that, to further our understanding of music's possible educational benefits, a few of the most promising mechanisms should be prioritized and studied more intensively.

It is important to note that any musical intervention will employ, at most, a few of the mechanisms covered in Table 1. For example, teaching students to sing a jingle might empha-size *mechanisms 1e* or *3a*, whereas helping students write their own song lyrics would likely emphasize a different mechanism such as *2d*. In the present study, we used the intervention of showing content-rich music videos so that we could efficiently gather input from hundreds of students spread over several different classes. The students' collective perceptions of the songs should be valuable in informing future song development and usage.

MATERIALS AND METHODS

Study Design

Our overall goal was to capture and classify undergraduate students' candid free responses regarding specific musical interventions. Exposing students to every possible type of musical intervention would have made the study overly complicated, so we focused on the common and easy-to-scale-up intervention of having students listen to educator-penned songs. Students were asked about songs that were typical of educator-written songs in that they covered key content via lyrics, but were modest in terms of musical sophistication. Students were asked about each song in an open-ended way so that their responses would not be biased by the researchers' preconceived notions (e.g., as represented by Table 1). That is, we wanted students to come up with their own reasons, rather than making selections from a list, so that the reasons given would more likely represent their genuine opinions.

Table 1. *Hypothesized mechanisms by which music could enhance STEM learning*

---

1. Promotes student engagement:
  1a. Grabs students' attention (1, 10)
  1b. Decreases students' stress levels (1, 26)
  1c. Delivers content via a familiar, welcome vehicle (2, 3)
  1d. Makes material less dry and more personal (11, 17, 34)
  1e. Builds classroom community via song-sharing, group singing, etc. (11, 13).
2. Promotes processing and analysis of content:
  2a. Helps teach the language of science (13)
  2b. Distills content into concise chunks (7)
  2c. Offers lyrics to analyze for deeper understanding (13)
  2d. Can be written by students as a form of "writing to learn" (9, 37)
3. Promotes retention and recall of content:
  3a. Uses rhymes and rhythms to function as mnemonic devices (9, 36)
  3b. Evokes motions to increase memorability (10, 13)
  3c. Activates multiple brain pathways to increase memorability (10, 13)

The study was conducted in two phases. *Phase 1*, the exploratory phase, gathered input on 5 songs from 440 students in 8 different courses at 2 universities. *Phase 2*, the validation phase, attempted to confirm the major conclusions of *phase 1* using 5 new songs and 53 new students.

Study Participants

Two medium-sized primarily undergraduate universities in the western United States participated in the study: The University of Washington Bothell (UWB) and San Francisco State University (SFSU). Both are considered "Masters Colleges and Universities: Larger Programs" by the Carnegie Classification of Institutions of Higher Education (14). Undergraduate students were enrolled in the study through nine different courses containing a major physiology component. These courses attracted a variety of students, including both nonscience and science majors, and included both lower- and upper-division offerings (Table 2).

Students were invited to participate in this study by their instructors toward the very end of the term (by which point which any song-related course material should have already been covered, if it was going to be covered at all). Students received a modest amount of extra credit for their participation, which consisted of taking a 20-min survey. Participation rates ranged from 86 to 100% (Table 2).

This study, as minimally invasive educational research, was judged exempt from Institutional Review Board review by both San Francisco State University and the University of Washington.

Songs Included in This Study

Of ~150 physiology-related songs written by the first author (6), 10 songs on mathematical aspects of physiology were selected for this study (Table 3). The mathematical physiology theme was pursued for two reasons. First, this study expanded on a previous study of songs on that topic (8), which, in turn, was based on the first and last authors' experience that students often struggle with mathematical aspects of physiology and could use extra help in this area. Second, we thought that using songs about equations would help students grasp the point of each song without a lot of additional explanation. The 10 selected songs represent varied musical styles, similar to the pop/rock genre. One song was a parody ("The Nernst Equation," based on "I Want You Back" by the Jackson 5); the other nine were original compositions. Each song was performed by the lead author (sometimes with friends) and audio-recorded as a simple "live" track with no overdubs or mixing. The recordings ranged from 17 to 54 s. A "video" was made for each song by adding a single PowerPoint slide displaying both the lyrics and the math equation covered.

Instructor ratings of song relevance. The instructor of each course rated each song's relevance to his/her course, according to the following 0–4 scale (Table 4):

- 0: Not closely related to the material covered in the course.
- 1: Not explicitly covered, but closely related to material covered in the course.
- 2: Explicitly covered during one session of the course (a lecture or laboratory).
- 3: Explicitly covered during multiple sessions of the course (lectures and/or laboratories).
- 4: One of the major themes of the course.

Since the instructors' relevance ratings and the students' viewing of the music videos were done toward the very end of the term, the instructors' relevance ratings should reflect students' actual exposure to the course content up to that point. For example, if an instructor rated a given song as a 3, it is safe to assume that the students had encountered the corresponding content multiple times before watching and reacting to that song.

Surveys

Phase 1 (exploratory phase) and *phase 2* (validation phase) each included an online survey administered via Google Forms. The two surveys were very similar. The core of each survey was a series of five web pages, each displaying an embedded music video (described above) and the following two questions:

- *Question 1*: "If your instructor showed you [*name of song*], would you use it as part of your studying for your current physiology

---

Table 2. *Study participants*

Course (Term)	Course Title	Course Type	School	Students Enrolled, *n*	Participants*, *n*
		Phase 1 (exploratory)			
BBIO 220 (Spring)	Introductory Biology (Animal & Plant Physiology)	Lower division, science majors	UWB	91	83 (28)
BBIO 220 (Summer)	Introductory Biology (Animal & Plant Physiology)	Lower division, science majors	UWB	36	31 (31)
BBIO 241†	Human Anatomy & Physiology I	Lower division, prenursing majors	UWB	28	27 (24)
BBIO 352†	Principles of Anatomy & Physiology II	Upper division, science majors	UWB	29	30‡ (27)
BIOL 100	Human Biology	Lower-division, non-science majors	SFSU	24	22 (21)
BIOL 212	Principles of Human Physiology	Lower division, prenursing majors	SFSU	44	41 (39)
BIOL 612	Human Physiology	Upper division, science majors	SFSU	138	123 (114)
BIOL 616	Cardiorespiratory Physiology	Upper division, science majors	SFSU	87	82 (76)
		Phase 2 (validation)			
BBIO 351†	Principles of Anatomy & Physiology I	Upper division, science majors	UWB	61	53 (47)

SFSU, San Francisco State University; UWB, University of Washington Bothell. *No. of students answering one or more survey questions (in parentheses is the number answering all survey questions). †Taught by the writer/performer of the songs. ‡One student must have completed the survey twice. We were unable to remove the duplicate because surveys were anonymous, and no two sets of answers were identical.

Table 3. *Songs assessed*

Title	Time, s	No. of Words*
Phase 1 (exploratory phase)		
"Fick's Law of Diffusion"	36	44
"The Nernst Equation"	36	51
"Surface Area-to-Volume Ratio"	37	48
"Poiseuille's Law of Laminar Flow"	41	61
"In-Lever, Out-Lever"	46	56
Phase 1 mean	39	52
Phase 1 median	37	51
Phase 2 (validation phase)		
"Pee Values"	39	33
"The Length Constant"	47	45
"Cardiac Output & Pulmonary Ventilation"	24	46
"The Goldman-Hodgkin-Katz Equation"	54	53
"Total Lung Capacity"	17	19
Phase 2 mean	36	39
Phase 2 median	39	45
Overall mean	38	46
Overall median	38	47

Links to all of these songs may be found at the following URL: https://faculty.washington.edu/crowther/Misc/Songs/2020songs.shtml. *Repeated lyrics were counted only once.

course?" [Choices: definitely, probably, maybe, probably not, definitely not]
- *Question 2*: "WHY would you be likely or unlikely to use [*name of song*]? (Please be honest; all reasons are valid. It's OK to give the same reasons for multiple songs.)" [blank text box]

The surveys were completed outside of class and, including the viewing of the music videos, took roughly 20 min for a student to complete. Songs were presented in the order shown in Table 3. (Whereas song order ideally would have been randomized, song order did not exert any obvious effects. For example, the last song of *phase 1* was the lowest-rated one, whereas the last song of *phase 2* was the highest-rated one.) Data were collected anonymously, after which students were given a link to a separate form so that they could register their names and receive extra credit for their participation from their instructor.

Coding of Students' Free Responses

In *phase 1* (exploratory phase), the first two authors independently read through all students' free responses on why they would be likely or unlikely to use each song in their studying. Among 12 common theme-identification strategies, we opted for "cutting and sorting" (29), in which each response is sorted into one or more categories.

Our set of categories was developed via multiple passes through the comments. Our starting point was the set of four musical categories

described and used by Ward et al. (37): Genre, Instrument, Lyrics, and Structure. However, we found that these options, in and of themselves, did not align well with students' comments. We omitted Genre (which was hardly ever mentioned by our students), replaced Structure with separate categories for Beat and for Melody (the structural aspects of songs commented on most frequently), and added categories of Catchiness, Learning, Length, Memory, Relevance, and Other to capture the many comments fitting these categories. Thus our final set of categories was as follows, with our operational definitions listed for each:

- Beat: The beat, rhythm, and tempo (speed) of the song.
- Catchiness: Songs' tendency to become lodged in the brain involuntarily or with little effort. Includes words like "catchy," "sticky," or "stuck in head." Does not cover remembering and memorizing per se, which reflect conscious, intentional effort and are considered to be in the Memory category.
- Instruments: Which instruments (including vocals) were used, and how well they were performed individually and/or as an ensemble.
- Learning: Gaining knowledge or understanding (as opposed to remembering, which is the Memory category). Includes different learning resources and styles. Comments about the lyrics (explanations, examples) rather than the song as a whole generally go in the Lyrics category.
- Length: The time elapsed, amount of information presented, and degree of repetition (if any).
- Lyrics: A song's words being clear or confusing, having helpful or distracting rhymes, having or lacking useful examples, etc.
- Melody: Whether the tune is pleasant to listen to, easy to sing, etc.
- Memory: Whether a song would help the student remember information (as opposed to the Learning category, which concerns learning or understanding information).
- Relevance: The song's applicability to the student and his/her studies, as well as comments that the material is so easy that there is no need for a song.
- Other: Features not covered by any of the above categories, sometimes because they are too broad or too cryptic to place elsewhere, including emotional reactions to a song as a whole, rather than a specific component.

Each categorized comment was further classified as positive (i.e., a reason why the song would be useful in studying), negative (i.e., a reason why the song would not be useful in studying), or mixed (i.e., a reason why the song was useful in one way and not useful in another, or of uncertain or context-dependent usefulness).

Examples of comments in each category are given in Table 5.

In *phase 2* (validation phase), the first author classified student responses using the same categories as in *phase 1*.

Table 4. *Instructors' ratings of songs' relevance to their courses*

	BIOL 616	BIOL 612	BBIO 352	BBIO 241	BBIO 220 (Spring)	BIOL 100	BIOL 212	BBIO 220 (Summer)	BBIO 351
Instructor	B	D	C	C	W	M	M	A	C
Fick	2	3	3	1	4	2	2	2	
Nernst	2	3	0	1	1	1	2	1	
SA/V	1	0	2	0	4	1	1	3	
Poiseuille	4	2	2	0	1	1	2	0	
Levers	0	0	4	2	0	0	1	0	
Pee									1
Length									3
CO&PV									1
GHK									3
TLC									1

As described in MATERIALS AND METHODS, the scale used was 0 (least relevant) to 4 (most relevant). Fick, "Fick's Law of Diffusion"; Nernst, "The Nernst Equation"; SA/V, "Surface Area-to-Volume Ratio"; Poiseuille, "Poiseuille's Law of Laminar Flow"; Levers, "In-Lever, Out-Lever"; Pee, "Pee Values"; Length, "The Length Constant"; CO&PV, "Cardiac Output & Pulmonary Ventilation"; GHK, "The Goldman-Hodgkin-Katz Equation"; TLC, "Total Lung Capacity."

Table 5. *Examples of coding of student responses*

Category	Positive Example(s)	Mixed Example(s)	Negative Example(s)
Beat	• "The beat to this song is catchy and is fun to listen to, so listening to this while studying would be helpful and would make studying not as dull."	• "The first part of this song was easy to follow, but the second [half] was a lot of information at a fast rate so I would have to read it instead of listen to it so I could get all the pieces."	• "The part of the song describing the formula was fast and wasn't easy to follow along with." • "I don't like the rhythm."
Catchiness	• "It is catchy and was instantly stuck in my head once the song ended." • "It helps concepts stick in my mind more, especially with a catchy tune."	• "The first part of the lyric is moderately catchy enough for me to remember. 'If you're small then it's high.' The rap at the end was not catchy enough for the concept to stick in my head."	• "Too uncatchy to stick in my mind." • "There is no catchiness in the song, it is too long and confusing for someone listening to it a couple of times."
Instruments	• "I think the voice is clear and is pretty. Song is easy to listen to." • "I like the audio from the guitar." • "The voice is great, its catchy and the way he is singing it is not annoying. A lot depends the way they are singing the song."	• "Compared to the previous song played, this has better music. I was initially interested in what I was hearing because it was a faster pace; however, when I heard the voice I no longer paid attention because it was slow and monotone and had to read the lyrics; even then I did not think the song had a good flow (because of the voice)."	• "The voice (way it's sung) for this song gives me a childish feeling. Like this song is meant for elementary or middle school students." • "The guitar playing in the background seems to be too loud, even louder than the guy singing the song, and with that, that creates further distraction in listening to the content to the song."
Learning	• "This song was very straight forward and easy to understand. I was able to know exactly what I need to do to apply the formula."	• "It is a little complicated but it might be worth looking over a couple times if you are having trouble with the equation."	• "I don't learn by listening to songs, I just need to write things out. I don't think this would help my learning." • "I am more of a visual learner."
Length	• "The song was to the point." • "The song is relatively short and simple with short verses, making it easy to memorize."	[No comments were rated mixed by both coders.]	• "The song is too long. I might use the first and second line but definitely not the last four." • "It's very long. I wouldn't try and sing this during a midterm in order to remember the formula."
Lyrics	• "I might use the song to study because the lyrics went smoothly with the beat of the song. That way it makes it easier to remember the lyrics." • "The lyrics include meaningful application of surface area, which is helpful."	• "Explains concepts but the actual equation in the second line isn't very smooth." • "I liked the beginning and the end, but the '6L to the 2 over L to the 3' was kind of confusing."	• "The rhyming does not help in this situation. The second part of the song is simply a paragraph that does not need to be sung." • "There is no explanation in the song."
Melody	• "Like the song because the melody is nice and addictive." • "The melody would possibly help students remember the words of the song."	• "Rap portion is too hard to remember because it is fast and does not have a unique melody, but melody section is memorable."	• "I do not like the melody of the song so I would never use it to learn something." • "I find the melody of the [song] to be distracting for the lyrics."
Memory	• "It is repetitive. It is easier to retain information when it is repeated." • "The equation itself is hard to remember so maybe using the song would help me remember it."	• This song seemed to be more effective because of the rhythm but the choppy ending makes it hard for me to use it as an actual remembering technique. I would say this song is good to remember the constant values used in the equation."	• "Just memorizing the equation would be simpler than learning the song." • "The material alone is a lot to remember; having them as lyrics with a melody makes it more difficult to retain because I have to remember how the 'song' goes."
Relevance	• "This equation can be used in all types of classes." • "I would use this because we learned about this on our past chapter."	[No comments were rated mixed by both coders.]	• "I'm a business major and have no need to remember this [unless] I was a Jeopardy contestant." • "We are usually given the Nernst equation on exams."
Other	• "This is what I'm talking about!!! I love it!" • "I honestly thought this was entertaining/funny."	• "I think the songs are better for the smaller equations. This song was a little more concept based but it worked for me."	• "Corny." • "The song is too odd to use." • "It doesn't help me."

Interrater Agreement on Coding of Responses in Phase 1

To quantify the agreement of the two researchers' coding of comments, we calculated Cohen's κ statistics (4). As shown in Table 6, agreement was substantial (κ > 0.6) for the categories of Beat, Catchiness, and Memory. Inclusion of words like "beat," "catchy," "memorize," and "remember" made comments relatively easy to sort into these categories. Other categories (e.g., Learning) were much broader, less associated with any particular words, and showed somewhat lower agreement between the two researchers (κ = 0.3–0.6). We considered this level of agreement to be an

Table 6. *Cohen's κ for assessing interrater agreement*

	Beat	Catchiness	Instruments	Learning	Length	Lyrics	Melody	Memory	Relevance	Other
Fick	0.79	0.82	0.82	0.39	0.58	0.56	0.26	0.58	0.27	0.31
Nernst	0.76	0.69	0.84	0.43	0.65	0.44	0.40	0.71	0.67	0.43
SA/V	0.87	0.85	0.67	0.41	0.28	0.49	0.28	0.75	0.35	0.33
Poiseuille	0.81	0.81	0.74	0.60	0.60	0.43	0.42	0.35	0.44	0.32
Levers	0.80	0.77	0.66	0.36	0.44	0.47	0.25	0.69	0.33	0.32
Arithmetic mean	0.81	0.79	0.75	0.44	0.48	0.48	0.31	0.69	0.41	0.34

According to Landis and Koch (16), κ values are often interpreted as follows: ≤0 = poor agreement, 0.01–0.20 = slight agreement, 0.21–0.40 = fair agreement, 0.41–0.60 = moderate agreement, 0.61–0.80 = substantial agreement, and 0.81–1.0 = almost perfect agreement. Fick, "Fick's Law of Diffusion"; Nernst, "The Nernst Equation"; SA/V, "Surface Area-to-Volume Ratio"; Poiseuille, "Poiseuille's Law of Laminar Flow"; Levers, "In-Lever, Out-Lever."

acceptable reflection of the challenges of categorizing diverse comments, and so we declined to iteratively discuss nonagreements and work toward stronger agreement, as is sometimes done in these situations. In particular, while κ for the Melody and Other categories did not reach the 0.41–0.60 range often considered to represent "moderate agreement" (16), these categories were used infrequently and were not a main focus of the analyses, so their relatively low κ values were deemed acceptable.

Mathematical Modeling of Students' Ratings

In *phase 1*, we planned to use multiple linear regression to determine the extent to which the dependent variable of students' ratings (of the usefulness of an educational song for their science learning) could be predicted from possible independent explanatory dichotomous variables: student sex (female or male), major (science major or nonscience major), whether the student was taught by the songwriter/performer (yes or no), whether student was okay with styles of all five songs collectively (yes or no), and the rating of the song's relevance by the student's instructor (0–4 scale, which we dichotomized into "related to the course" or "not related to the course"). A full regression model was ultimately not created, as discussed below (*Results of Mathematical Modeling of Students' Ratings*).

RESULTS

Students' Ratings of Songs' Usefulness

The students collectively perceived differences in the usefulness of the five *phase 1* songs. As summarized in Fig. 1, the distributions of ratings were significantly different from song to song, according to a χ^2 test of homogeneity ($\chi^2 = 55.34$, df = 16, $P < 0.00001$); that is, we reject the null hypothesis that the distributions of ratings are the same for all five songs. In addition, and, for the purposes of this paper, more importantly, each song provoked a wide range of student opinions. Each song got the highest possible rating ("definitely") from at

least 10% of students and the lowest possible rating ("definitely not") from at least 10% of students.

The variety of students' reactions to the songs (Fig. 1) does not itself indicate whether each student tended to have a similar opinion of all of the songs (i.e., whether a student tended to find them all useful or find none of them useful). As it turned out, only 16% of the students (64 of 395) gave the same usefulness rating to all five songs, whereas 46% (180 out of 395) rated at least one song as "probably" or "definitely" and at least one song as "probably not" or "definitely not." These percentages (16% and 46%) suggest that most students did not simply accept or dismiss the entire batch of songs as a group, but instead paid attention to each individual song and considered the possible merits of each.

Results of Mathematical Modeling of Students' Ratings

For each of the five songs, we first checked for multicollinearity among the potential independent variables, finding that there was only one correlation whose magnitude exceeded 0.55: a correlation (for the "Levers" song) of 0.906 between "student's professor deems song relevant to course" and "student's professor is the singer/songwriter." For each variable, however, an examination of both dot plots (for the 0 and 1 levels) revealed essentially similar distributions, making clear that that variable would not be a meaningful predictor, and so a multiple regression model was not viable for this data set.

Reasons Given by Students for Rating Songs as Useful or Not Useful

The main purpose of this study was not to gather usefulness ratings for these particular songs, which are just a few of the
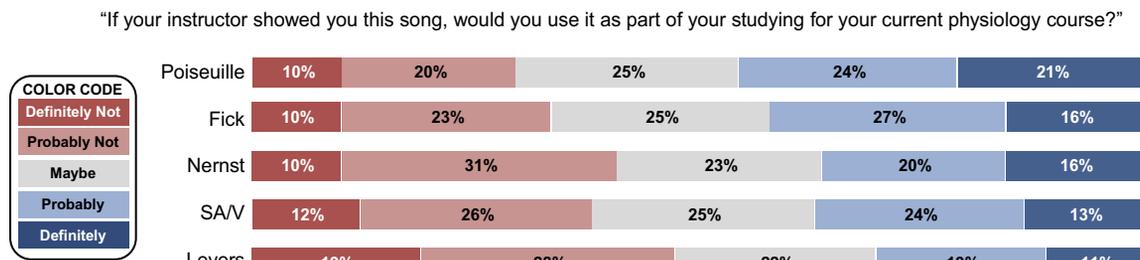

Fig. 1. Students' responses to the question, "If your instructor showed you this song, would you use it as part of your studying for your current physiology course?" The songs listed are the *phase 1* songs introduced in Table 3. Only responses from the students who rated all five songs (*n* = 395) are included. Poiseuille, "Poiseuille's Law of Laminar Flow"; Fick, "Fick's Law of Diffusion"; Nernst, "The Nernst Equation"; SA/V, "Surface Area-to-Volume Ratio"; Levers, "In-Lever, Out-Lever."

Table 7. *Counts of students' comments, with subclassifications as positive, mixed, negative, or no consensus on subclassification*

	Beat	Catchiness	Instruments	Learning	Length	Lyrics	Melody	Memory	Relevance	Other	(Catchiness + Memory)/Total
Fick	22	123	36	40	13	33	2	168	4	19	245/405
P	12	82	2	13	4	20	0	96	0	4	
M	0	2	0	0	0	0	0	0	0	0	
N	10	27	32	21	8	8	2	44	4	12	
NC	0	12	2	6	1	5	2	28	0	3	
Nernst	39	78	32	47	23	17	12	129	14	26	183/385
P	26	47	12	18	3	7	10	57	0	11	
M	1	1	3	0	0	0	0	3	0	5	
N	7	14	15	25	17	9	2	53	12	6	
NC	5	16	2	4	3	1	0	16	2	4	
SA/V	50	70	18	65	7	36	3	114	8	17	160/394
P	9	51	6	23	1	14	1	57	1	8	
M	2	1	0	2	0	2	1	2	0	4	
N	34	13	11	35	4	16	1	45	6	4	
NC	5	5	1	5	2	4	0	10	1	1	
Poiseuille	45	84	15	56	18	40	5	126	6	21	174/378
P	18	69	3	17	5	21	2	80	0	8	
M	0	0	0	2	0	1	0	8	0	3	
N	21	5	10	31	10	10	2	27	4	6	
NC	6	10	2	6	3	8	1	11	2	4	
Levers	37	59	22	52	13	26	5	98	11	23	147/371
P	13	28	0	22	1	10	3	41	1	7	
M	1	0	0	0	0	0	0	0	0	2	
N	21	27	20	26	12	13	1	43	9	10	
NC	2	4	2	4	0	3	1	14	1	4	
Totals											909/1,933

This *rightmost* column shows two numbers: first, a count of all of the comments that were coded by both researchers as being in the Catchiness category and/or the Memory category, and, second, the total number of comments received for that song. Numbers for Catchiness and Memory do not sum to the (Catchiness + Memory) totals because some comments were placed in both categories, but are counted only once in the (Catchiness + Memory) total. M, mixed; N, negative; NC, no consensus on subclassification; P, positive. Fick, "Fick's Law of Diffusion"; Nernst, "The Nernst Equation"; SA/V, "Surface Area-to-Volume Ratio"; Poiseuille, "Poiseuille's Law of Laminar Flow"; Levers, "In-Lever, Out-Lever."

thousands of existing educational STEM songs (7), but to explore the reasons why individuals might or might not find each song useful. To this end, we collected their free responses on the usefulness of the songs (median length 18 words for *phase 1* and 16 words for *phase 2*), and we classified these comments into 10 thematic categories, as described above. Table 7 shows the distribution of the comments into the categories for the five *phase 1* songs. For every song, the most used category was Memory, and the second most used category was Catchiness. In all, 47% of *phase 1* comments, nearly one-half, were classified into the Memory category, the Catchiness category, or both.

Aside from showing the overall distribution of comments among the 10 categories, Table 7 also shows each category's subcategories of positive, mixed, negative, and no consensus counts. The totals of these subcategories corroborate the indication by Fig. 1 that these relatively simple, nonprofessional songs elicited widely varying reactions among students. For example, in the Memory category, 57 students explicitly said that "The Nernst Equation" song would be helpful in remembering the Nernst equation, whereas 53 others explicitly said that the song would not be helpful in this way. Examples of positive, negative, and mixed comments in the Memory category are shown in Table 5. Similarly, divergent reactions, although with lower numbers, were seen for the categories of Beat, Learning, and Lyrics. In contrast, most comments in the Catchiness category were considered positive, and most in the Instru-ments, Length, and Relevance categories were considered negative.

Song length. The songs used in *phase 1* were all between 36 and 46 s long, in line with previous recommendations to keep them short (8, 17). For the 14 comments classified by both coders as being positive regarding song length, all 14 praised the song's conciseness. For the 51 comments classified by both coders as being negative about song length, 50 criticized the song as being too long, whereas only 1 said that the song was too short. Thus most students who raised the issue of length felt that the songs were appropriately short or could be even shorter.

Song relevance. A song's relevance to the curriculum should also affect its perceived usefulness by the students. Only a few students explicitly raised this issue in their comments (Relevance column of Table 7). Furthermore, our attempt to model a relationship between relevance and perceived usefulness did not reveal a noticeable trend (see above). However, the anticipated trend, songs rated as highly relevant by the instructor tended to be rated more useful by students, can be discerned in two "extreme" cases, where a given song had very high relevance in one course and very low relevance in another course. Table 8 shows the two pairs of courses where two songs had opposite relevance for the two courses. For both of these pairs of courses, instructors' ratings of the songs' relevance corresponded with students' ratings of the songs' usefulness. Collectively, these data suggest that a song's relevance has a demonstrable but not overwhelming impact on its perceived usefulness.

Table 8. *Pairwise comparisons of songs and courses show that relevance correlates with perceived usefulness*

Courses	Songs	Ratings Summary
BBIO 352 (UWB) vs. BIOL 612 (SFSU)	Levers vs. Nernst	BBIO 352: Instructor rated Levers much more relevant than Nernst (4 vs. 0). Students rated Levers more useful than Nernst (4.9 vs. 4.0).
		BIOL 612: Instructor rated Nernst much more relevant than Levers (3 vs. 0). Students rated Nernst more useful than Levers (4.0 vs. 3.4).
BBIO 220 (UWB/summer) vs. BIOL 616 (SFSU)	Poiseuille vs. SA/V	BBIO 220: Instructor rated SA/V much more relevant than Poiseuille (3 vs. 0). Students rated SA/V more useful than Poiseuille (3.4 vs. 3.0).
		BIOL 616: Instructor rated Poiseuille much more relevant than SA/V (4 vs. 1). Students rated Poiseuille more useful than SA/V (3.3 vs. 2.9).

For this table, students' usefulness ratings were converted to a 1-to-5 scale, as follows: definitely not = 1, probably not = 2, maybe = 3, probably = 4, definitely = 5. All between-song differences in students' usefulness ratings were statistically significant ($P < 0.01$), according to unpaired two-tailed t tests. SFSU, San Francisco State University; UWB, University of Washington Bothell. Levers, "In-Lever, Out-Lever"; Nernst, "The Nernst Equation"; Poiseuille, "Poiseuille's Law of Laminar Flow"; SA/V, "Surface Area-to-Volume Ratio."

Phase 2: Validation of Selected Phase 1 Results

By the end of *phase 1*, our working hypotheses were that the songs' perceived usefulness hinges especially on their length and their ability to aid memorization of content. To see whether these emergent hypotheses could be supported further, we examined a new set of comments collected from 53 new students on 5 new songs. For these *phase 2* students, we found that remembering content was once again the dominant issue, with 53% of *phase 2* comments falling into the Catchiness and/or Memory categories (range: 40% for the "Total Lung Capacity" song to 63% for "The Length Constant" song).

Phase 2 also allowed us to further explore the issue of song length, since the new songs varied considerably in their lengths (17–54 s; 19–53 words), in contrast to the *phase 1* songs' relatively uniform lengths (36–46 s; 44–61 words). We gave each *phase 2* song a "song length approval rating" by subtracting the number of negative length-related comments from the number of positive length-related comments. There was a strong negative correlation between song length approval rating and song length, regardless of whether song length was expressed as the duration in seconds ($R^2 = 0.84$) or the number of words ($R^2 = 0.81$) or the sum of the two ($R^2 = 0.96$; Fig. 2). In other words, the shorter a song was, the more highly its length was rated. Moreover, the shortest of the 10 songs
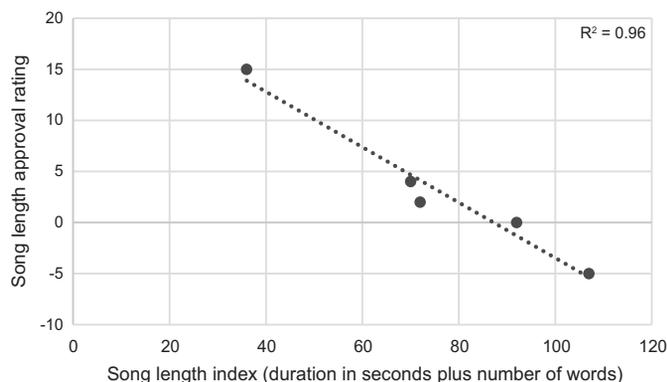

Fig. 2. In *phase 2*, the lengths of short songs were viewed more positively than the lengths of longer songs. Since students may perceive song length based on both the song duration in seconds and the number of words delivered, we summed these two quantities to get a "song length index" for each song (*x*-axis). We defined "song length approval rating" (*y*-axis) as the number of positive comments about the song's length minus the number of negative comments about the song's length.

("Total Lung Capacity") had the highest usefulness rating (3.9 on a 5-point scale, calculated as described in the legend of Table 8).

DISCUSSION

In this study, we gained insight into what undergraduate students value in educational songs. For the 10 songs studied here, the students collectively perceived the songs' usefulness to depend on both academic factors (e.g., clarity of words and relevance to the course) and aesthetic values (e.g., rhythm and singing quality). Aspects that students perceived as especially important were brevity and usefulness as mnemonic devices. These aspects are probably related, since, all else being equal, shorter songs are easier to memorize.

This study had four notable limitations. First, our intervention was designed for research rather than for pedagogical effectiveness. Just as students do not learn optimally from passively listening to lectures (12), they are unlikely to learn optimally from passively watching science music videos, as they did here. To spur additional learning, students should actively interact with the songs, e.g., by performing content-rich movements to illustrate the lyrics and/or by answering study questions about the songs (13, 25). Therefore, the first author has supplemented this study's songs, and many others, with resources, such as study questions, sheet music, and karaoke tracks. These resources were not advertised to our study subjects at the time, but are now available to all interested teachers and students (6).

A second limitation is that the nature of our study probably skewed students' comments toward some themes and away from others. We showed the students online music videos without facilitating group singing or group discussions, which probably emphasized the music's fact-delivery aspects (*mechanisms 1c, 2a, 2b,* and *3a* of Table 1) over its sociocultural aspects (*mechanism 1e* of Table 1). The exact phrasing of our questions (i.e., "If your instructor showed you [*name of song*], would you use it as part of your studying for your current physiology course?") may have also nudged students' answers toward the realm of absorbing and retaining facts. Nevertheless, the fact that so many students independently reacted to the songs with memory-related comments suggests that they generally associate educational songs with memorization. This focus on memorization may reflect widespread beliefs that music can be an effective mnemonic device (10) and/or

that memorization is a central task of science classes in general (22).

A third limitation is that, in this exploratory study, we did not profile individual students extensively enough to capture all of the factors that may have influenced their perceptions of the songs. The dot plots showed no significant influence of a student's sex or major on their responses; however, we did not ask about other possibly relevant factors (e.g., grade point average, ethnicity, general views on science education, personal experience with music performance). Future research could attempt a more comprehensive accounting of such influences on students' perceptions, perhaps through more extensive surveys or interviews of a smaller number of students.

A final study limitation is that we did not formally assess learning: we only collected data on student perceptions of learning. The usual caveats regarding student feedback certainly apply here, e.g., students are not always the best judges of their own learning (15). Nevertheless, as suggested by the recent surge of research on student perceptions (24, 27, 31), an expanded understanding of such perceptions may ultimately lead to better teaching and learning. In this case, our findings on students' concerns about memory and song length have several implications for STEM instructors and educational researchers.

Regarding memory, STEM instructors should be aware that, if they incorporate music into a course, many students will assume (fairly or not) that the main point of the music is to help them memorize information. Instructors should not feel bound by this assumption, but should address it by explicitly informing students of any important goals beyond memorization per se. Deemphasizing memorization, if appropriate, may in turn alter students' musical preferences, e.g., students may tolerate longer songs better if they understand that they are not expected to memorize them.

Regarding song length, our evidence that students prefer songs to be as short as possible, perhaps <30 s, corroborates and extends previous data on this point (8), but seems to be in conflict with the lengths of most educational STEM songs used in high school and college. In a national curated database of 178 songs about statistics (19, 21, 28), the median song length was found to be 1 min 55 s, i.e., over twice as long as the longest song in the present study, whereas in a database of 191 physics songs (32), the median song length was 2 min 7 s (Fig. 3). The finding that most existing database songs greatly exceed students' preferred lengths holds up, regardless of whether one looks at the databases' parody songs or original tunes or both. Considering only database songs that (like 9 of the 10 songs in this study) had original tunes, the median length is 2 min 28 s for the statistics songs and 2 min 12 s for the physics songs.

As with the memory issue, students' song length preferences should inform songwriters' and instructors' work without necessarily dictating it. For example, those who create educational songs should at least consider whether a one-verse jingle would suit their purposes. This perspective is relevant to those who create song parodies, since an effective parody need not last as long as the original song (17). Among other benefits, keeping songs brief may help them be experienced as pleasant interludes rather than as distracting digressions and should mini-
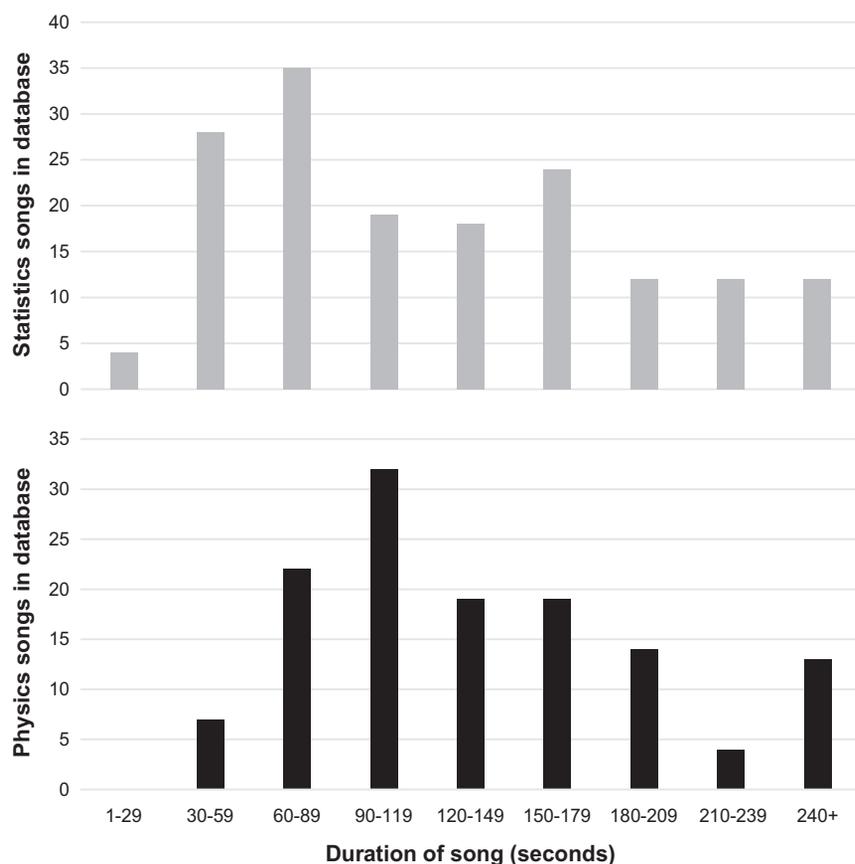

Fig. 3. Distribution of song lengths in two discipline-specific databases of educational STEM songs hosted at CAUSEweb.org (Ref. 28; *top*) and PhysicsSongs.org (Ref. 32; *bottom*). Bars do not sum to the total number of songs in the database because audio recordings were unavailable for some songs.

mize any backlash from students who prefer not to learn from songs.

ACKNOWLEDGMENTS

DISCLOSURES

AUTHOR CONTRIBUTIONS

REFERENCES

1. **Allen R, Wood WW.** *The Rock 'n Roll Classroom*. Thousand Oaks, CA: Corwin, 2013.

2. **Bennett A.** *Popular Music and Youth Culture: Music, Identity, and Place*. London: Palgrave Macmillan, 2000.

3. **Bennett A.** Subcultures or neo-tribes? Rethinking the relationship between youth, style and musical taste. *Sociology* 33: 599–618, 1999.

4. **Cohen J.** Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 70: 213–220, 1968. doi:10.1037/h0026256.

5. **Crowther G.** Using science songs to enhance learning: an interdisciplinary approach. *CBE Life Sci Educ* 11: 26–30, 2012. doi:10.1187/cbe.11-08-0068.

6. **Crowther GJ.** *STEM Songs* (Online). https://faculty.washington.edu/crowther/Misc/Songs/ [20 Dec 2019].

7. **Crowther GJ.** The SingAboutScience.org database: an educational resource for instructors and students. *Biochem Mol Biol Educ* 40: 19–22, 2012. doi:10.1002/bmb.20567.

8. **Crowther GJ, Davis K, Jenkins LD, Breckler JL.** Integration of math jingles into physiology courses. *J Math Educ* 8: 56–73, 2015.

9. **Crowther GJ, Ma AJ, Breckler JL.** Songwriting to learn: can students learn A&P by writing content-rich lyrics? *HAPS Educator* 21: 119–123, 2017. doi:10.21692/haps.2017.025.

10. **Crowther GJ, Williamson JL, Buckland HT, Cunningham SL.** Making material more memorable ... with music. *Am Biol Teach* 75: 713–714, 2013. doi:10.1525/abt.2013.75.9.16.

11. **Emdin C, Adjapong E** (Editors). *Hip-hop as Education, Philosophy, and Practice. #HipHopEd: The Compilation on Hip-hop Education*. Rotterdam, the Netherlands: Brill-Sense, 2018, vol. 1.

12. **Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP.** Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci USA* 111: 8410–8415, 2014. doi:10.1073/pnas.1319030111.

13. **Governor D, Hall J, Jackson D.** Teaching and learning science through song: exploring the experiences of students and teachers. *Int J Sci Educ* 35: 3117–3140, 2013. doi:10.1080/09500693.2012.690542.

14. **Indiana University School of Education. Center for Postsecondary Research.** *The Carnegie Classification of Institutions of Higher Education. 2018 Update Public File* (Online). https://carnegieclassifications.iu.edu/downloads.php [23 Dec 2019].

15. **Kornell N, Bjork RA.** Learning concepts and categories: is spacing the "enemy of induction"? *Psychol Sci* 19: 585–592, 2008. doi:10.1111/j.1467-9280.2008.02127.x.

16. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174, 1977. doi:10.2307/2529310.

17. **Lesser LM.** Mathematical lyrics: noteworthy endeavours in education. *J Math Arts* 8: 46–53, 2014. doi:10.1080/17513472.2014.950833.

18. **Lesser LM.** Third VOICES Conference on Teaching STEM with Music, September 22–23, 2019. *J Humanist Math* 9: 372–373, 2019. doi:10.5642/jhummath.201902.32.

19. **Lesser LM, Pearl DK.** Functional fun in statistics teaching: resources, research and recommendations. *J Stat Educ* 16: 1–11, 2008. doi:10.1080/10691898.2008.11889572.

20. **Lesser LM, Pearl DK, Weber JJ III.** Assessing fun items' effectiveness in increasing learning of college introductory statistics students: results of a randomized experiment. *J Stat Educ* 24: 54–62, 2016. doi:10.1080/10691898.2016.1190190.

21. **Lesser LM, Pearl DK, Weber JJ III, Dousa DM, Carey RP, Haddad SJ.** Developing interactive educational songs for introductory statistics. *J Stat Educ* 27: 238–252, 2019. doi:10.1080/10691898.2019.1677533.

22. **Li M, Zheng C, Liang JC, Zhang Y, Tsai CC.** Conceptions, self-regulation, and strategies of learning science among Chinese high school students. *Int J Sci Math Educ* 16: 69–87, 2018. doi:10.1007/s10763-016-9766-2.

23. **Maeda J.** STEM + art = STEAM. *STEAM J* 1:34,2013.

24. **Malau-Aduli BS, Alele FO, Heggarty P, Teague PA, Sen Gupta T, Hays R.** Perceived clinical relevance and retention of basic sciences across the medical education continuum. *Adv Physiol Educ* 43: 293–299, 2019. doi:10.1152/advan.00012.2019.

25. **Mangan JM, Newman D, Doss KB, Virani SN.** Improving science content learning with choreographed songs at an astronomy summer camp. *Int J Sci Educ* 9: 101–113, 2019. doi:10.1080/21548455.2019.1571257.

26. **Modell HI, DeMiero FG, Rose L.** In pursuit of a holistic learning environment: the impact of music in the medical physiology classroom. *Adv Physiol Educ* 33: 37–45, 2009. doi:10.1152/advan.90149.2008.

27. **Olson J, Rinehart J, Spiegel JJ, Al-Nakkash L.** Student perception on the integration of simulation experiences into human physiology curricula. *Adv Physiol Educ* 43: 332–338, 2019. doi:10.1152/advan.00202.2018.

28. **Pearl DK.** *CAUSEweb Fun Collection* (Online). https://www.causeweb.org/cause/resources/fun/ [23 Dec 2019].

29. **Ryan GW, Bernard HR.** Techniques to identify themes. *Field Methods* 15: 85–109, 2003. doi:10.1177/1525822X02239569.

30. **Skorton D, Bear A** (Editors). *The Integration of the Humanities and Arts with Sciences, Engineering, and Medicine in Higher Education: Branches from the Same Tree*. Washington, DC: National Academies Press, 2018.

31. **Slominski T, Grindberg S, Momsen J.** Physiology is hard: a replication study of students' perceived learning difficulties. *Adv Physiol Educ* 43: 121–127, 2019. doi:10.1152/advan.00040.2018.

32. **Smith WF.** *PhysicsSongs.org* (Online). http://ww3.haverford.edu/physics-astro/songs/ [20 Dec 2019].

33. **Smolinski K.** Learning science using music. *Sci Scope* 35: 42–45, 2011.

34. **Stovall D.** We can relate hip-hop culture, critical pedagogy, and the secondary classroom. *Urban Educ* 41: 585–602, 2006. doi:10.1177/0042085906292513.

35. **Tanner KD.** Order matters: using the 5E model to align teaching with how people learn. *CBE Life Sci Educ* 9: 159–164, 2010. doi:10.1187/cbe.10-06-0082.

36. **Thaut MH, Peterson DA, McIntosh GC, Hoemberg V.** Music mnemonics aid verbal memory and induce learning-related brain plasticity in multiple sclerosis. *Front Hum Neurosci* 8: 395, 2014. doi:10.3389/fnhum.2014.00395.

37. **Ward SJ, Price RM, Davis K, Crowther GJ.** Songwriting to learn: how high school science fair participants use music to communicate personally relevant scientific concepts. *Int J Sci Educ* 8: 307–324, 2018. doi:10.1080/21548455.2018.1492758.