

First results from the Dark Skies, Bright Kids astronomy club draw-a-scientist test

Christian R. Hayes¹,* Allison M. Matthews¹, Yiqing Song¹, Sean T. Linden, Robert F. Wilson,
Molly Finn¹, Xiaoshan Huang, Kelsey E. Johnson, Anne M. McAlister,
Brian Prager, and Richard Seifert

University of Virginia, Post Office Box 400325, Charlottesville, Virginia 22904-4325, USA

Sandra E. Liss²

Department of Physics, Radford University, Radford, Virginia 24142, USA

Andrew M. Burkhardt³

*Center for Astrophysics—Harvard & Smithsonian,
60 Garden Street, Cambridge, Massachusetts 02138, USA*

Nicholas Troup

*Department of Physics, Salisbury University,
1101 Camden Avenue, Salisbury, Maryland 21804-6860, USA*

 (Received 30 August 2019; accepted 14 April 2020; published 28 May 2020)

This paper presents the first results from draw-a-scientist tests (DASTs) over five years that were used to measure the effect of 8–10 week long astronomy clubs and week long summer camps on 3 rd–5 th grade elementary school students’ perceptions of scientists. We facilitated these DASTs prior to these clubs or camps, which provide a baseline for a student’s initial conception of scientists, and once at the end, to determine whether their conception changed, possibly as a result of their involvement. In total we analyze 89 pairs of DASTs using a numerical grading scheme designed to quantify the presence of various features in the drawn scientist and their activities. We find that there is a gender imbalance in both the pre- and postclub drawings, with only 32% and 35%, respectively, of students drawing female scientists. We also find that a third to a half of the scientists have a stereotypical appearance and/or are performing stereotypical activities. Although we find insignificant changes (<5%) in most categories, we do find an 8% increase in the number of scientists that have a stereotypical appearance, which is worth following up, but a significant 12% decrease in the number of scientists who are performing stereotypical activities. In addition, we present some possible improvements to implementing DASTs and discuss other possible assessments that could provide a more direct method of gauging the effect of these astronomy clubs or camps.

DOI: [10.1103/PhysRevPhysEducRes.16.010131](https://doi.org/10.1103/PhysRevPhysEducRes.16.010131)

I. INTRODUCTION

Dark Skies, Bright Kids (DSBK), founded in 2009, is a University of Virginia-based outreach organization that offers sustained out-of-school time (OST) experiences for underserved students in Virginia. The goals of DSBK are to (i) enhance upper elementary student interest in science, (ii) encourage scientific inquiry and engagement, and (iii) teach basic astronomical concepts, through

fun and hands-on activities. In DSBK clubs, volunteers guide the students through hands-on learning experiences designed to stimulate curiosity and enhance scientific understanding using astronomy topics from local elementary-school curricula as a “gateway” into other astronomy topics not covered in elementary school curricula [1,2]. Most volunteers are astronomy graduate students who (with a changing, but roughly time-averaged 50-50 female-male gender balance), in addition to teaching basic astronomy concepts, encourage the student participants to both revise stereotypical ideas of who can be a scientist and discover that science is an exciting and fun process [2]. The hope is that by engaging students in science we can help them see that anyone (including themselves) can be a scientist.

Internal assessments of the DSBK after school program indicate that student perceptions of who scientists are and

*crh7gs@virginia.edu, dsbk-assessments@virginia.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

what scientists do [3] change over the course of the DSBK after school program, which could be due, in part, to the influence of this program. Here we propose to formalize these internal findings by using the draw-a-scientist test (DAST) [4], which was previously developed to investigate students' perception of scientists, and which we intend to use to investigate (i) the nature of upper elementary school students' preconceived notions of science and scientists and (ii) whether these can be broadened through an extended educational intervention in the form of a DSBK astronomy club or summer camp. Typical work on student perceptions is undertaken in a traditional classroom setting [5–9]. However, our research was focused on the effects of nontraditional science enrichment programs provided by external organizations and typically led by professional scientists. Such programs are provided by numerous organizations, and most are supported as “broader impact” activities within larger science, technology, engineering, and mathematics (STEM) research grants, but the DAST has not been used extensively to assess the impact these types of programs have on student perceptions of scientists.

Interest alone is often not enough to get youth involved in STEM fields. In a report examining career interest, many students reported high career aspirations, but only a small percentage reported seeing themselves as scientists [10]. Not only was STEM consideration lower than many other career aspirations, it was disproportionately low when compared to the number of students reported to have an interest in science. The discrepancy stemmed from how they perceived the scientific community. Students reported seeing professional scientists as “brainy,” and if this was not part of their self-concept, they could not see themselves as scientists in the future. Race and class were also barriers. Males, Asians, and those with high levels of cultural capital reported seeing themselves as scientists. Girls, especially those with low cultural capital, were much less likely to see themselves as scientists. These factors were amplified in the case of Black students. The problem as illustrated by this report was that the students had a hard time envisioning themselves as scientists. Even if they expressed an interest in science, they could not see themselves in their mental representation of what a scientist is: brainy, White, and male.

Research by Bhattacharyya *et al.* [11] found that after participating in a week-long science summer camp African-American students reported having more positive attitudes towards science and wanting to take more science classes, and the percent of students who saw themselves majoring in science more than doubled. Another study found that girls' participation in an OST science camp that emphasized active learning and science practices helped the participants see themselves as scientists and dispel the notion that science is primarily a male pursuit [12].

To measure a change in students' perception of the scientific community, DSBK adopted the DAST. The initial purpose of the DAST [4] was to learn at what age the

well-known stereotypical image of the scientist first appears in children. It has now been used in a variety of settings to show the persistence and pervasiveness of this image, even in preservice teachers [8]. Students who participate in DSBK were asked to complete a DAST twice, drawing what they think a scientist looks like and describing what a scientist does—as part of the application process, and as part of a post-club survey. During the application process, which is facilitated by a school liaison, parent, or guardian, potential participants were asked to complete the DAST [5] (an example DAST is given in the Appendix). The students completed this initial DAST before meeting any members of DSBK to avoid potential bias of what they may think the volunteers want them to draw, or by the appearance of the volunteers themselves. The postclub survey was designed to collect similar data on student perception by asking them to repeat what they have done in their application. In this way we were able to construct a dataset for each student in a club and use previously developed DAST evaluation metrics to measure the change in their perception of who scientists are and what they do, over the course of our 8–10 week club, or week-long summer camp.

II. DATA COLLECTION AND GRADING

A. Data collection

The data used for this study were collected over five years from 2013 to 2017 from both our after school astronomy clubs that meet once a week for 8–10 weeks during the school semester, and our week-long astronomy summer camps that meet daily from 8 AM to 3 PM (Monday–Friday). Prior to each semester club and summer camp we sent out applications directly to either an administrator at a school or, in the case of the summer camps, a local point of contact who could distribute applications to schools and students in the surrounding areas. Each application contained a DAST along with logistical details, which included parent contact information, a media release form, and in some cases a short explanation from the students about why they wanted to be in our astronomy club. Further, for students who elected to provide us with demographic information, we recorded their race, gender, and status as recipients of free or reduced lunch. Because of the inconsistent collection of these demographics, an investigation into correlations between DAST results and various student demographics will be analyzed in future work.

The DASTs within the initial applications represent the students' conception of what a scientist looks like before ever interacting with DSBK members or our astronomy programs (hereafter “before” DASTs). These before DASTs are important for establishing a baseline for how our students generally perceive scientists, so that we can compare with their perceptions of scientists after they

attend our clubs or summer camps. One goal of our program is to engage students in science and let them see that anyone (including themselves) can be a scientist. A main goal of our study is to see if and how the students' images of scientists changed after interacting with DSBK members and each other in our astronomy program. However, since no members of DSBK are present when students are performing their before DAST, we do not always know the conditions they were performed in. Parents or teachers may be assisting students with ideas of what to draw, such as famous scientists (e.g., Albert Einstein, Neil DeGrasse Tyson, etc.). Similarly, students may be biased or encouraged to draw scientists looking at stars or planets, since they are applying to an astronomy program. Overall, while there are some aspects of the before DAST that we cannot control, our method aims for the greatest minimization of outside perspectives and opinions which can alter a student's perspective about what a scientist looks like.

To evaluate whether the DSBK program has at least a short term effect on how students view scientists, we perform a second DAST at the end of the program (hereafter the "after" DAST). These after DASTs are administered by DSBK volunteers during the last day of the semester clubs or summer camps, when all students are gathered in the same room and given ~10–15 min to work on their drawings. The DSBK volunteers are specifically instructed not to tell the students what to draw. However, the proximity of the volunteers may have some influence on what the students draw (for example, students will occasionally draw specific DSBK volunteers for their after DAST). In this way, volunteer demographics may impact students' drawings, and ideally we would want to investigate if students' drawings reflect the demographics of DSBK volunteers. In general, the DSBK club volunteers are typically Caucasian with a small representation from ethnic or racial minorities, are balanced in terms of gender, and come from a range of socioeconomic status. However, since the DSBK volunteers vary significantly day to day and change from one club to the next (as volunteers join or leave), tracking the impact of volunteer demographics cannot be done in a detailed way.

Similarly, the environment that the students perform their after DAST may influence what they draw. In some cases, students may be distracted by each other and draw scientists doing activities similar to something that occurred that day. Additionally, because the students are doing their DASTs while still in the Astronomy club or camp, they may be thinking about astronomy related ideas and have a tendency to draw more astronomy-related scientists. Furthermore, the short amount of time spent on the DASTs during the last day of club may limit the amount of detail the students add to their drawings; while there were no time constraints on the before DASTs, however, performing the DASTs on the last day of club

or camp assures a higher completion rate by students than if a postclub or camp survey were sent by mail.

While we try to perform a before and after DAST for every student, occasionally, students sign up or attend the astronomy clubs or camps without having completed an application, so they will not have a recorded before DAST, or some students will leave part of the way through a club or camp or miss the final day and will never complete an after DAST. In this paper, we are interested in tracking the effects that attending our club or camp has on the students in the short term, so we only compare before and after DASTs for students that have completed both. This normalizes any preconceptions or ideas unique to an individual student in both sets of DASTs. Over the four years and nine astronomy clubs or camps that are analyzed in this paper (with ~12–20 students per club or camp), 132 students completed a DAST, 89 students completed both a before and after DAST. This set of 89 pairs of DASTs along with our grading scheme, which is discussed below, provides the dataset for the remainder of this paper.

B. "Grading" the DASTs

To perform quantitative analysis for the drawings collected, instead of qualitatively describing trends, each DAST was decomposed into 42 representative symbols (listed with descriptions in Table I), which summarize the different features that may be used to describe the scientists presented in the DAST drawings as modified from Mason *et al.* [13]. For example, a typical drawing of a scientist might be a white male wearing a lab coat, safety goggles, and/or working with glassware on a lab bench. We want to track the type of scientists reflected in DASTs taken before and after our astronomy clubs and camps. To do so, we record the qualities of the drawn scientists by counting the occurrence of features such as "male," "lab coat," "glassware," "lab bench," as well as qualities that reflect astronomy-related activities and studies, such as "telescopes," "stars or planets," etc. These symbols are designed to provide a fairly comprehensive description of both the characteristics of the drawn figure (gender, demeanor, and appearances) and their working environments (equipment and setting). In general, we adopted a binary grading system as a way to quantify how frequently a specific symbol is present in the drawings. We assigned "1" if symbols were expressed in the drawing or description and "0" if they were not expressed.

To illustrate a typical grading process, we use Fig. 1 (drawn by one of the authors) as an example. In this drawing, we clearly see the presence of a telescope, a computer, a pencil, and a piece of paper lying beneath the drawn figure, as well as stars (objects in space) in the background. The figure drawn wears a pair of glasses or goggles, and seems to be dressed in a lab coat while obviously holding a piece of glassware, which is further confirmed through the text ("...doing experiment...a tube

TABLE I. DAST grading rubric.

Grading symbol	Description
Personal characteristic	Physical characteristics of the scientists
Lab coat	Scientist is wearing a lab coat
Eyeglasses	Scientist is wearing glasses
Facial hair	Scientist has facial hair
Pencils or pens	Scientist has pencils or pens
Unkempt	Scientist's appearance is unkempt
Research symbols	Common research related items
Glassware	Glassware is present
Microscope	A microscope is present
Bunsen burner	A bunsen burner is present
Plants or animals	There are plants or animals present
Paper or clip board	There is paper or a clip board present
Other (nonastronomy)	There is another (nonastronomy related) research related item present
Astronomy symbols	Astronomy related items
Telescope or planetarium	A telescope or planetarium is present
Objects in space	An object in space is present
Rockets	There is a rocket present
Other astronomy object	There is another astronomy related object present
Knowledge symbols	Objects related to knowledge or thought
Books	Books are present
Cabinets	Cabinets are present
Lab Bench	A lab bench is present
Speech or thought bubbles	The scientist has a speech or thought bubble
Technology symbols	Objects related to technology
Computer or calculator	A computer or calculator is present
Machines	A machine of some kind is present
Gender	The gender of the scientist, identified based on physical characteristics and names if present (only one of the following is reported for each scientist in a DAST)
Man	The scientist is distinctly male (e.g., common male name, has facial hair)
Woman	The scientist is distinctly female (e.g., common female name, has long hair or eyelashes)
Indeterminate (likely male)	The scientist does not have an easily distinguished gender, but has more stereotypically male associated qualities (e.g., short or no hair)

(Table continued)

TABLE I. (Continued)

Grading symbol	Description
Indeterminate (likely female)	The scientist does not have an easily distinguished gender, but has more stereotypically female associated qualities (e.g., wearing a dress)
Race or ethnicity	The race or ethnicity of the scientist, identified by drawn skin color, names, or if the scientist is famous (only one of the following is reported for each scientist in a DAST)
Caucasian	Scientist is Caucasian
African-American	Scientist is African-American
Hispanic	Scientist is Hispanic
Native American	Scientist is Native American
Asian	Scientist is Asian
Not colored in	Scientist is not colored in and a race or ethnicity cannot be determined
Other (not human colored)	Scientist is not human colored (e.g., the scientist may be an alien or a nonhuman animal)
Other (B&W scan)	The DAST was scanned in black and white and a race or ethnicity cannot be ascertained
Demeanor	Apparent demeanor of the scientist
Happy	The scientist appears happy
Sad	The scientist appears sad
Neutral	The scientist appears to have a neutral mood
Appearance	Appearance of the scientist and their actions
Eccentric	The scientist appears to be eccentric (e.g., crazy hair, blood shot eyes, unkempt)
Sinister	The scientist appears to be sinister or performing experiments that reflect negatively on scientists (e.g., making bombs, experimenting on animals crying for help, making monsters)
Neutral	Scientist is not doing broader than traditional activities or negative ones. Appears neutral or working in a traditional environment (e.g., a chemistry lab, making potions, not doing anything, etc.)
Broader than traditional	Scientist is working in a non traditional setting or using unusual or outdoor lab equipment (e.g., working in a garden, using a telescope, working out in space)
Stereotypical	Records of if the scientist appears to be stereotypical or involved in stereotypical activities
Activity	The scientist is performing a stereotypical activity (e.g., mixing chemicals, launching a rocket, experimenting on people or animals)
Appearance	The scientist has stereotypical appearance (e.g., wearing a lab coat or Albert Einstein-like)

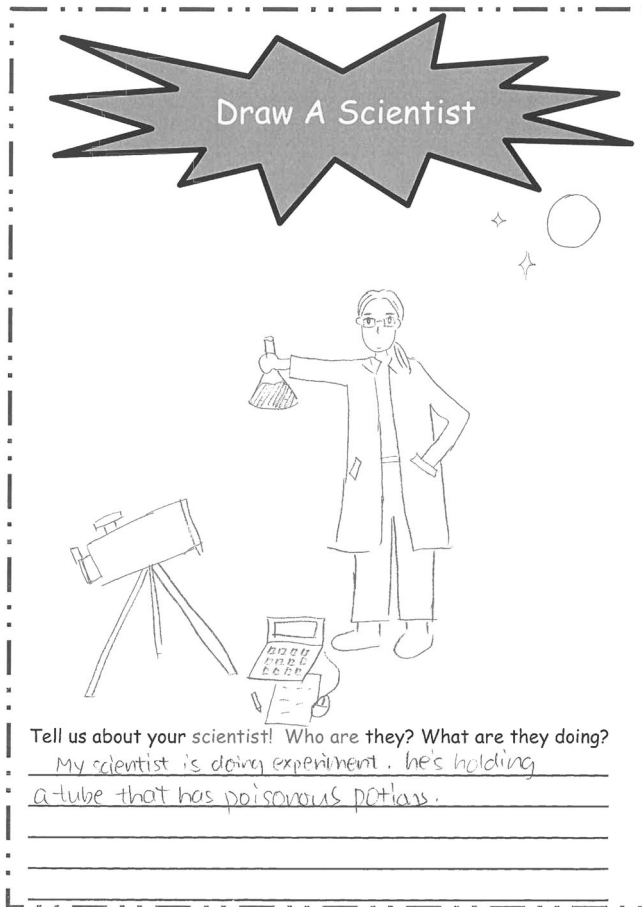


FIG. 1. Example DAST drawing by the authors to illustrate some of the features common to student-drawn DASTs.

that has poisonous potion”). By the drawing alone, the gender and race of the figure is hard to determine, especially given that the drawing is not colored in, but the use of “he” in the text description confirmed the figure to be male. Since the figure was not given a smile nor a frown, we determine the demeanor to be neutral. Despite the presence of a poisonous potion which has sinister implications and astronomy symbols which implies a broader than traditional working environment, the portrayed action does not appear overly negative nor broader than traditional, therefore the appearance is graded as neutral. As for stereotypical activity and appearance, even though astronomy symbols are presented in the drawing, a man conducting chemical experiment in goggles and lab coat is still a stereotypical impression of scientific research and scientists. Even though the stereotypical activity or appearances were graded based on general impressions, inherited biases of such grading methodology are discussed at the end of this section as well as in Sec. III.

To account for the fact that grading results can vary from one grader to another, each DAST was independently graded by two unique graders and differences in their grading results were later settled through group discussion.

Groups were typically composed of 4–6 independent graders, and any conflicting symbol scores were reviewed and decided as a collective. The subject whose DAST was being graded was made anonymous to the graders by removing any indication of a name, and attempting to choose graders who were not present at the club that could potentially recognize a particular subject’s work. However, this was not always possible, and graders of the DAST would occasionally be the same as those who administered the DASTs.

III. RESULTS AND DISCUSSION

A. Correlations

Quantifying the amount of correlation between features present in students’ drawings illustrates trends present in our DAST sample, and additionally allows us to better understand underlying assumptions in the evaluation metrics we adopted. We computed the Pearson correlation coefficient for each pair of the DAST symbols in our metric after combining both the before and after datasets (Fig. 2). The correlation statistic ranges from -1 to 1, from a strong anticorrelation to a strong correlation, respectively. Pairs showing either a strong correlation or anticorrelation may simply reflect the definition of the symbols (e.g., “woman” anticorrelates with “man”), the methodology (e.g., glassware correlates with stereotypical activity) or actual underlying biases held by our sample of elementary students. We excluded 9 demographic symbols representing race, ethnicity, and socioeconomic status from the calculation due to lack of records, leaving 33 symbols (1056 distinct pairings) in the correlation matrix.

Out of the 528 unique symbol pairings, 27 show positive correlation with a coefficient of 0.25 and above, and 11 show negative correlation with a coefficient of -0.25 and below. These symbols that are more strongly correlated or anticorrelated ($|\text{Correlation Index}| > 0.25$) are listed in Table II. However, many of these correlation or anticorrelation values do not necessarily reflect actual trends present in the data collected, but rather they reflect intrinsic relevance or irrelevance between the symbols concerned. For example, pairs such as “paper or clipboard” and “pens or pencils,” “cabinets” and “books,” lab bench and glassware, etc., are often complementary, therefore showing strong positive correlations. Pairs such as woman and man, “broader than traditional appearance” and “unkempt,” “neutral demeanor” and “happy demeanor” are mutually exclusive, thus showing strong anticorrelation.

Meanwhile, some correlations or anticorrelations reflect inherent biases in the grading methodology we adopted. For example, glassware, lab bench and neutral appearance are associated with a stereotypical activity or appearance presumably because these are typically associated with the stereotypical idea of a scientist. The same line of reasoning also produces the anticorrelations between stereotypical

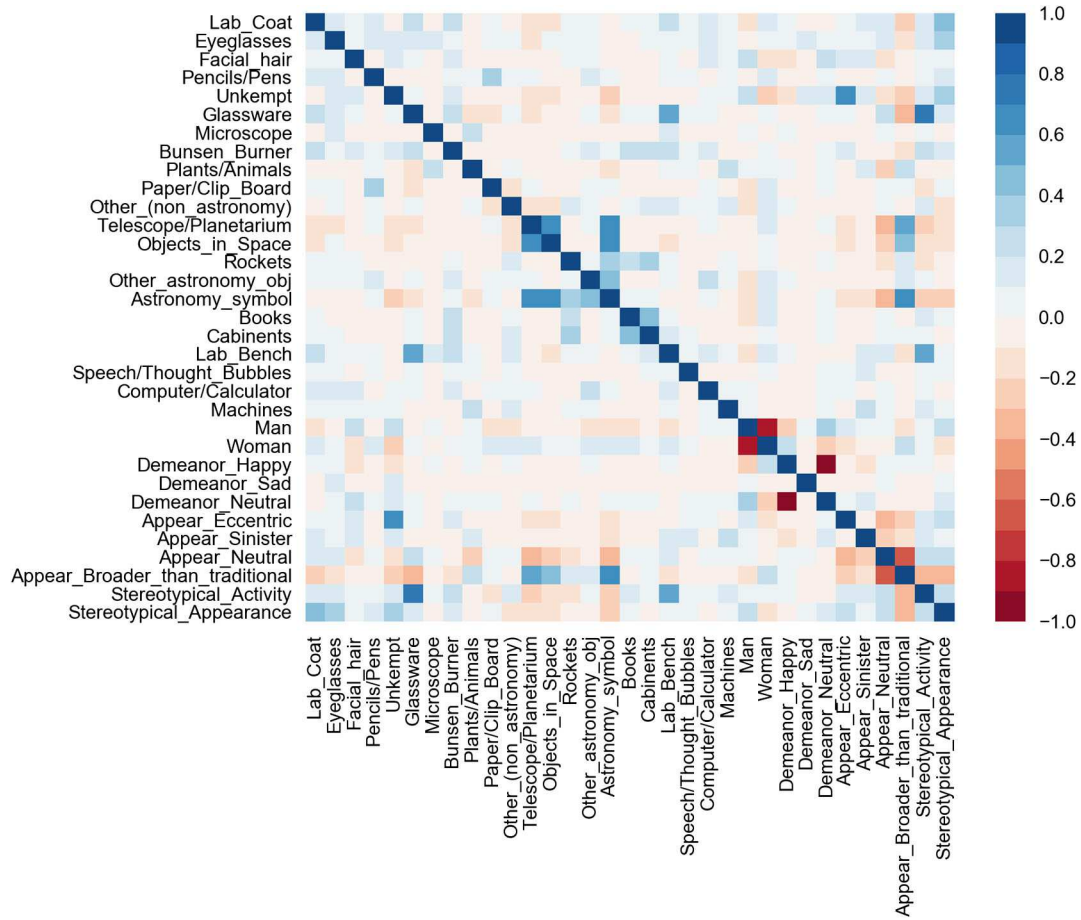


FIG. 2. Matrix showing the correlation index for each pair of DAST symbols. All indices fall between -1 and 1, where -1 indicates complete anticorrelation and 1 complete correlation.

TABLE II. Correlation indices for pairs of symbols with correlation > 0.25 or correlation < -0.25.

DAST symbols		Correlation	DAST symbols		Correlation
Glassware	Stereotypical activity	0.71	Rockets	Cabinets	0.37
Telescope or planetarium	Astronomy symbol	0.67	Pencils or pens	Paper or clip board	0.35
Objects in space	Astronomy symbol	0.65	Rockets	Astronomy symbol	0.32
Telescope or planetarium	Objects in Space	0.65	Unkempt	Stereotypical appearance	0.32
Astronomy symbol	Broader than traditional	0.62	Man	Neutral demeanor	0.30
Unkempt	Eccentric appearance	0.62	Neutral appearance	Stereotypical activity	0.29
Glassware	Lab Bench	0.57	Bunsen burner	Lab bench	0.29
Telescope or planetarium	Broader than traditional	0.56	Woman	Happy demeanor	0.28
Lab bench	Stereotypical activity	0.55	Glassware	Bunsen burner	0.28
Lab coat	Stereotypical appearance	0.48	Plants or animals	Broader than traditional	0.26
Other astronomy object	Astronomy symbol	0.44	Bunsen burner	Books	0.26
Objects in space	Broader than traditional	0.42	Glassware	Neutral appearance	0.26
Books	Cabinets	0.40	Stereotypical activity	Stereotypical appearance	0.25
Eyeglasses	Stereotypical appearance	0.40			
Happy demeanor	Neutral demeanor	-0.94	Astronomy symbol	Stereotypical activity	-0.26
Man	Woman	-0.89	Woman	Neutral demeanor	-0.26
Neutral appearance	Broader than traditional	-0.68	Unkempt	Broader than traditional	-0.29
Eccentric appearance	Neutral appearance	-0.39	Glassware	Broader than traditional	-0.30
Broader than Traditional	Stereotypical activity	-0.38	Broader than traditional	Stereotypical appearance	-0.34
Astronomy Symbol	Neutral appearance	-0.38			

activity and all astronomy related symbols such as telescope or planetariums, happy demeanor, and broader than traditional appearance.

Despite the expected outcomes described above, a few pairs of symbols highlight interesting perspectives held by our sample of elementary-school students. Based on our analysis, man has strong association with unkempt while woman shows the opposite. Similarly, woman correlates with happy demeanor more than man does. This is likely reflecting the underlying social influences on the subjects throughout daily life. It is shown in a series of studies revealing the early influence of gender stereotypes on children’s perceptions on “brilliance” [14]. For children as young as 6 years old, the notion of brilliance is already more strongly associated with male while “niceness” is strongly associated with female. From the children’s perspectives, women or girls tend to be modest, friendly and approachable, but not necessarily smart compared to men or boys. Given this phenomenon, it is then not surprising that in our sample we observe overall happier-looking portrayal of female scientists compared to male scientists, which likely reflect the social emphasis of “niceness=women” that children are subjected to in their daily lives [14].

B. Analysis of the before DAST sample

Our before DASTs serve as a general snapshot for how rising 3rd–6th grade students (ages 8–11) view scientists. When discussing student perceptions of scientists, one of the central issues is determining at what age children develop and form opinions about who scientists are (race, gender, etc.), and what they do. The study of Mead and Métraux [15], which established the canonical definition of a stereotypical scientist as “an older male in a white lab coat,” highlights the fact that by the time students reach high school they have already formed opinions about scientists. Formal science instruction is often sporadic in early elementary school and often does not occur regularly until late elementary to middle school [16]. During this time, children likely form gender stereotypes about science and mathematics based, in part, on the teachers’ attitudes and beliefs, as some correlational studies have suggested [17,18]. The age demographic of our students is uniquely

suitable to catch this stereotype formation in action by comparing the gender ratio and presence of stereotypes in the drawings from our DAST sample with those of previous DAST analyses that examined students over a larger age range.

In the landmark study of “draw-a-scientist” tests, Chambers [4] analyzed the drawings of nearly 5000 elementary-school students, administered from 1966 to 1977 primarily in the United States and Canada. Of the 5000 students, only 28 (0.6%) drew female scientists, of which none were drawn by male students. Miller *et al.* [19] performed a meta-analysis of nearly five decades of DASTs from ~20000 students and found that the percentage of female scientists drawn grew to 28% on average in more recent studies (1985–2016). Excluding the data from this later period, Miller found only a weak trend that the percentage of female scientists drawn does increase over time. Including the Chambers’ study in the analysis, however, Miller found this historical time effect to be statistically significant in all regression models, concluding that children have drawn more female scientists since the 1960s and 1970s.

Table III lists the percentage of male and female scientists drawn in our before dataset. We observe that 32% of students draw female scientists before participating in our astronomy club. This is in agreement with Miller *et al.* [19] and strengthens the statistical significance of the historical trend of the percentage of female scientists drawn. Although women are still largely underrepresented in several STEM disciplines (particularly computer science, physics, and astronomy), the representation of women as scientists in the media has grown since the time of Chambers’ study. The children’s magazine *Highlights* features general content to encourage “curiosity, thinking ability, and imagination” but devotes a sizable portion to science and science-related fictional stories. Previs [20] analyzed the frequency of females versus males in *Highlights* science stories from 1947–2010 and found a significant change over time. Since the 1960s, the mention of females to males has increased steadily, and the average percentage of mentions of women in the magazine has consistently been greater than the true percentage of women in science at that time. This increased representation

TABLE III. Before and after club statistics, presenting the number of before and after DASTs exhibiting each of the selected symbols (and the corresponding percentage in parentheses). The Anderson-Darling k -sample significance is given for each before-after pair, measuring the percent likelihood that the before and after measurements are drawn from the same population, with a cap of >25% indicating which pairs of before and after samples are consistent with no change.

	Men	Women	Broader than traditional	Stereotypical appearance	Stereotypical activity	Astronomy symbols
Before	53 (68%)	25 (32%)	20 (26%)	28 (36%)	34 (44%)	19 (24%)
After	51 (65%)	27 (35%)	21 (27%)	34 (44%)	25 (32%)	22 (28%)
AD significance	> 25%	> 25%	> 25%	> 25%	7%	> 25%

of women in media may help children associate both men and women with scientists and contribute to the larger percentage of women being drawn in DASTs.

Furthermore, the fraction of “stereotypical” scientists drawn by the students in our sample (either in appearance or activity) is nearly equal to the that of male scientists drawn. This indicates that many of our students share the Mead and Métraux [15] prescription of who scientists are, and that this bias is already well established by late elementary school. The similarity between the numbers of male and stereotypical scientists drawn reflects a strong gender stereotype of scientists as male and influences the fractional participation of women in science. This is in agreement with earlier studies which found that a larger proportion of women in science courses and careers correlates with a weaker association of science with men over women [21].

We see an anticorrelation between broader than traditional and stereotypical appearance, but a positive correlation between broader than traditional appearance and astronomy symbols. The relationships among these variables can be explained through our bias towards classifying a DAST as broader than traditional if it reflects a nontraditional view of scientists working in a nontraditional setting (i.e., not in a lab). The percentage of students in our sample who drew “astronomy symbols” was significantly higher than a similar indicator of a nontraditional science environment used by Finson [22]: “scientist doing work outdoors.” From a sample of 94 total students, he found that only 3% of them drew scientists doing work outdoors, a number far less than the 36% of our sample who drew scientists in an astronomy setting (i.e., with a telescope). We posit that this discrepancy is likely due to the circumstances in which the before DASTs were given; as a part of an application for inclusion in an astronomy after school or summer program. We do not believe the higher percentages of broader than traditional and astronomy symbols we observed are an accurate indicator of student perception of the scientific working environment.

C. Before and after comparison

For the following analysis, we considered only students for which we had a DAST from both before and after our clubs or camps. We also only included DASTs which showed a single scientist (rather than a group) to simplify the analysis. Since the number of students who drew several scientists in a DAST is small compared to the total number of DASTs (11 out of 89), this should not influence the results of the analysis. This leaves us with 78 complete before and after pairs that show a single scientist.

This total includes both the DASTs from summer camps and semester-long clubs. Different formats for the week-long programs versus the weekly after school programs could have different impacts on the students’ after DASTs,

so differences between these two subsamples is worth investigating. Comparing the percent occurrence of symbols between the summer camp and semester club after DASTs, the small number of DASTs from the summer camps or semester clubs separately (39 for each) results in large standard deviations on these occurrence percentages.

Almost all of the differences between the summer camp and semester club after DASTs are within $\sim 1\sigma$ of each other (according to binomial statistics). Excluding the symbols related to how the DASTs were scanned [“Not Colored In” versus “Other (B&W scan)”] and after combining “indeterminate likely male or female” with their respective genders, we find only two symbols that have a larger, $\sim 2\sigma$ difference, “Telescope or planetarium” and “Caucasian” (most commonly appearing when students draw a famous scientist, whose race or ethnicity can be readily determined). These two symbols show up in 6 and 7 more DASTs after the summer camps, respectively, than after semester clubs, and may indicate some difference in the impact of these two programs.

However, since these differences correspond to few actual DASTs, we cannot rule out that they are due to differences between the individual students attending the separate events. Therefore, given the relatively low significance of the differences between the summer camps and semester clubs with the present sample, we have chosen to combine these two samples to improve our statistical assessment of OST experiences on students’ perception of scientists in general in the present work. We leave these possible programmatic differences to be examined in a future work when larger samples of DASTs have been performed and analyzed by DSBK.

To understand the change of students’ perception of scientists over the DSBK Astronomy clubs or camps, we compared the following characteristics of the DAST drawn before and after the club: male, female, broader than traditional, stereotypical appearance, stereotypical activity, and astronomy. For the male and female comparisons, we also include the “indeterminate likely male” and “indeterminate likely female” in the male and female sample, respectively. For the astronomy comparison, we recorded a score of 1 if the DAST included any astronomy symbols, such as telescopes, rockets, stars, planets, and other astronomical objects, while a score of 0 indicated that the students’ drawing did not contain any of these symbols, implementing our binary grading scheme as explained in Sec. II A.

We show these comparisons in Fig. 3 and provide the number of DASTs expressing each symbol (with the corresponding percentage) in Table III. Additionally we perform an Anderson-Darling (AD) k -sample test to statistically measure the similarity between our before and after DASTs. For each symbol that we examine, we show the AD “significance” in Table III, i.e., the percent likelihood that the number of DASTs displaying that

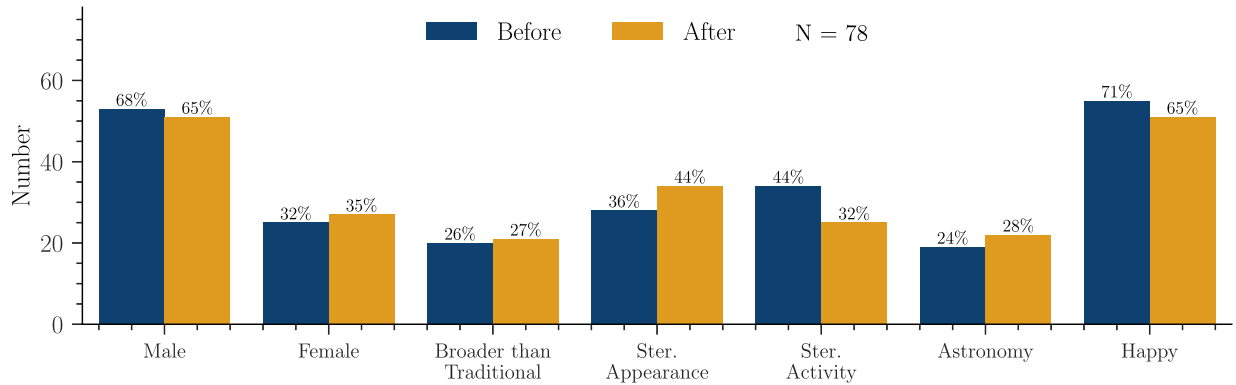


FIG. 3. Percentage of DASTs exhibiting the following symbols in our DAST grading rubric: a male scientist, a female scientist, a broader than traditional appearance of a scientist (i.e., a scientist in a nontraditional setting), a scientist with a stereotypical appearance, a scientist performing a stereotypical activity, or the presence of an astronomy related feature. These percentages are shown for both the before set of DASTs (blue) and the after set of DASTs (gold).

symbol before and after come from the same distribution. An AD significance $>25\%$ indicates that the before and after samples are consistent with being drawn from the same distribution for that symbol.

Within this grading scheme, we do measure small changes in the above-mentioned categories, but as seen in Table III, these are often statistically insignificant. We see only a slight decrease in the percentage (-3%) of male scientists drawn and, correspondingly, an increase in the percentage of female scientists drawn of $+3\%$ in the after sample. This may be due to the DSBK astronomy club aim to instruct students both verbally and by example that anyone, regardless of race, gender, ethnicity, socioeconomic status, etc., can become a scientist. In doing so, students (particularly female students) could begin to visualize and draw a scientist like themselves after having been in the club. However, these differences are not at a significant level given our small sample size, and motivate future increases to our DAST sample. Similarly, we observe a slight increase in the number of astronomy-themed scientists in the DASTs by $+4\%$, which may be influenced by students' attendance at an astronomy-specific science camp, but requires further scrutiny.

The largest differences we find among the above symbols are in both stereotypical activity and stereotypical appearance between the before and after DAST samples. While the DASTs characterized as stereotypical in appearance grew by 8% , it too is not significant by the AD k -sample test (AD significance $> 25\%$). However, the stereotypical activity symbol fell by 12% between the before and after DASTs, and, when performing an AD test, is significant at the 7% level, meaning that there is only a 7% chance these two values were drawn from the same distribution. While this is only a moderate statistical significance, it may indicate that the participation of students in DSBK Astronomy camp or clubs had an impact on students' perception of what scientists do, and could be a result of the students' interaction with non-traditional

scientific equipment and experiments (i.e., using telescopes or participating in cratering demonstrations, instead of experiments where one mixes chemicals). As with the results from our other symbols, we eagerly anticipate analyzing a larger collection of DASTs to further strengthen or clarify the significance of this change.

D. Possible improvements

The purpose of our DASTs is twofold, (i) to get a baseline for the general conception elementary students have of scientists, and (ii) to use the DAST as a measure of the effectiveness of the changing students' perceptions of scientists over the course of DSBK astronomy clubs and camps in the hope that they have been encouraged to envision themselves (or others around them) as scientists by their involvement in the program. Our results show little change in the student perception of scientists over the course of a DSBK club or camp, besides perhaps a slight decrease in the number of stereotypical appearance and activities drawn. However, our "baseline" as measured by our before DASTs is not without bias, and this makes it more difficult to establish a baseline perception of scientists and then to identify the changes during the astronomy clubs or camps.

One example of this is the slight decrease in the number of astronomy objects and activities being shown. There is an expectation that, by knowing that they are applying for an astronomy club or camp, students will be more likely to draw astronomy related objects and activities on the before DAST when applying to the club or camp. However, to verify this, we would need a comparison sample in which students do not know they are applying for an astronomy related event.

Other difficulties that arise in analyzing DASTs are in ensuring uniformity in the level of oversight, and (although it is minor) uniformity in the level of detail from students in their before and after DASTs. In terms of uniformity of

oversight, students are not in the same environment for their before and after DASTs. When drawing their before DASTs in the application, they may be influenced by their parents, guardians, or teachers, who administer the DAST with very little instruction from DSBK. When drawing their after DASTs on the last day of club or camp, they may be influenced by other students or their environment, despite the ability of DSBK members to minimize distractions. These factors potentially add bias when we try to determine what the students think scientists look like, or who could be a scientist. Additionally, the different settings for the before and after DASTs may mean that students have different lengths of time to draw and add detail to their DASTs. Typically the after DASTs take place in a more time-limited setting. This could affect their final products and how we analyze the drawings.

There are several ways that we could improve our methods to create a more standardized version of our DAST and better assess how students are possibly affected by our semester clubs and summer camps. It would be ideal to have a more controlled and uniform environment for students to work on their DAST. One possibility would be to separate the before DASTs from the applications, so that while the students are working on their DAST they are unaware that it will be for an astronomy program. Another way to improve uniformity would be to have the same conditions or environment for the students to do their before and after DASTs, and to perform the after DASTs at a later date rather than on the last day of the club (when the students have astronomy topics actively on their mind) to gauge the longer term evolution of students' perception of scientists following a DSBK program.

In terms of grading and analysis, we plan to update our grading strategy and reanalyze current DASTs to make note of when students draw themselves or DSBK members. This will serve as an indicator of the frequency at which they can recognize specific individuals in their lives as scientists rather than fictional or historical characters. In the future, we also plan to combine the DASTs with student demographics to examine how students of different gender, race or ethnicity, or socioeconomic status view scientists and whether DSBK has possibly impacted this view. As more DSBK clubs and camps are completed, we also intend to look for trends in student preconceptions over time either due to societal changes or due to the changes in the DSBK program as volunteers and the activities cycle through the years.

Outside of possible improvements to the current DASTs, we may be able to implement other evaluations of the DSBK's impact on students that more directly address our interests. For example, a survey or questionnaire could be given a few weeks or months after a DSBK club or camp has been concluded, with questions asking about what students got out of the club or camp, and how much of it they recall. Over a longer period of time DSBK could also

follow up with students at different stages in their education to track their interests and progress, or to again assess how much they remember from the club or camp that they attended (either in terms of material retention or their experience). Performing an evaluation over a longer baseline would help to assess the long term impact of DSBK, which is currently unaddressed by the DASTs.

IV. CONCLUSION

Here we present the first analysis of DASTs from five years of astronomy clubs that meet once a week for 8–10 weeks and from week-long astronomy camps both run by the Dark Skies, Bright Kids outreach program and serving about 12–20 3rd–5th grade elementary school students per club or camp. The purpose of these DASTs is to gauge students' initial conceptions of scientists and to assess change in how students think about science and who can be a scientist over the course of DSBK programs. To do this, we have presented students with a DAST prior to each DSBK program and at the culmination of each program. In this paper we analyze the first 4 years of this DAST assessment program, for which we have DASTs from a total of 132 students, 89 of which have both a complete set of before and after DASTs, allowing us to track changes over the course of their attendance in a DSBK club or camp.

We have analyzed these DASTs by quantifying their features according to the binary grading rubric given in Table I. The resulting aggregate statistics show that there is a general gender bias in the scientists both before and after the DSBK clubs or camps, with only 32% of students drawing female scientists prior to the DSBK program, with a small, insignificant increase of 3% following the program. This is fairly consistent with the recent results found by Miller *et al.* [19] that show the percentage of female scientists drawn by elementary school students has risen over the past 50 years or so, but still only sits around 28% in the most recent studies (up to 2016). This relatively low percentage comes despite a near gender balance in most of the clubs or camps, or even female dominated populations in some cases. This seems to reflect known patterns, that young students more often think of men as scientists and that women are underrepresented in STEM fields. On top of this underrepresentation of female scientists, we found that the female scientists that were drawn tended to have a happy demeanor, which may be another example that even at these ages children subscribe to the stereotype that associates niceness with women, but brilliance with men [14].

Additionally, we find that over half of students initially drew stereotypical scientists (e.g., wearing lab coats, or being disheveled or unkempt) and were doing stereotypical activities, like working in a chemistry lab, mixing chemicals, or performing a similar type of experiment. We also found that about one-third of all of the students initially

drew their scientist interacting with an object related to astronomy, however, we believe that this is inspired by their application to an astronomy club or summer camp, which they are aware of prior to their before DASTs. Comparing these baselines with the scientists drawn at the end of the clubs or camps, we found little change in the presence of astronomical objects, but did find that the presence of stereotypical science activities decreased by 12%, significantly different to the 7% likelihood level (i.e., there is only a 7% chance that the before and after observations are consistent with each other). This may indicate some success in opening the minds of the students to the range of possibilities of what scientists can do.

In the future we plan to divide these aggregate results by the demographics of the students who performed these DASTs and plan to reanalyze the DASTs to determine whether students were drawing themselves or DSBK members. With these additions, we hope to assess whether underrepresented minorities in STEM fields change their

conceptions of scientists over the course of the DSBK programs, and to see if they begin to envision themselves or people they know (such as the DSBK members) as scientists. This analysis can help the DSBK program further assess its impact on the community and general public, and could be combined with other future assessments that help determine this impact, such as long-term follow-up. Through these analyses we also hope to illustrate the effectiveness of different assessments for other STEM volunteer outreach programs.

ACKNOWLEDGMENTS

The authors thank the anonymous referees for their supportive and constructive feedback. The authors would also like to acknowledge the support from NSF Grants No. AST-1413231 and No. AST-1716335 and the David and Lucile Packard Foundation that have made this research possible.

-
- [1] B. Prager *et al.*, Dark Skies, Bright Kids! Year 5. American Astronomical Society, AAS Meeting No. 223, No. 444.04, 2014.
 - [2] L. E. Bittle *et al.*, Dark Skies, Bright Kids! Year 7. American Astronomical Society, AAS Meeting No. 227, No. 248.07, 2016.
 - [3] S. E. Liss *et al.*, Dark Skies, Bright Kids! Year 6. American Astronomical Society, AAS Meeting No. 225, No. 243.07, 2015.
 - [4] D. W. Chambers, Stereotypic images of the scientist: The draw a scientist test, *Sci. Educ.* **67**, 255 (1983).
 - [5] C. R. Barman, K. L. Ostlund, C. C. Gatto, and M. Halferty, Fifth grade students' perceptions about scientists and how they study and use science, AETS Conf. Proc. 688 (1997).
 - [6] M. Wyer, Intending to stay: Images of scientists, attitudes toward women, and gender as influences on persistence among science and engineering majors, *J. Women Minorities Sci. Engin.* **9**, 1 (2003).
 - [7] C. R. Barman, Students' views of scientists and science: Results from a national study, *Science and Children* **35**, 18 (1997).
 - [8] K. D. Finson, J. B. Beaver, and B. L. Cramond, Development and field test of a checklist for the draw-a-scientist test, *School Sci. Math.* **95**, 195 (1995).
 - [9] W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010101 (2006).
 - [10] K. L. Archer, J. DeWitt, J. F. Osborne, J. S. Dillon, B. Wong, and B. Willis, *ASPIRES Report: Young people's science and career aspirations, age 10–14* (King's College London, London, 2013), p. 40.
 - [11] S. Bhattacharyya, T. P. Mead, and R. Nathaniel, The influence of science summer camp on African American high school students' career choices, *School Sci. Math.* **111**, 345 (2011).
 - [12] K. Riedinger and A. Taylor, "I Could See Myself as a Scientist": The potential of out-of-school time programs to influence girls' identities in science, *Afterschool Matt.* **23**, 1 (2016).
 - [13] C. L. Mason, J. B. Kahle, and A. L. Gardner, Draw-A-Scientist Test: Future implications, *School Sci. Math.* **91**, 193 (1991).
 - [14] L. Bian, S.-J. Leslie, and A. Cimpian, Gender stereotypes about intellectual ability emerge early and influence children's interests, *Science* **355**, 389 (2017).
 - [15] M. Mead and R. Métraux, Image of the scientist among high-school students: A pilot study, *Science* **126**, 384 (1957).
 - [16] P. S. Smith, E. R. Banilower, K. C. McMahon, and I. R. Weiss, *The national survey of science and mathematics education: Trends from 1977 to 2000* (Horizon Research, Inc., Chapel Hill, NC, 2002).
 - [17] C. Keller, Effect of teachers' stereotyping on students' stereotyping of mathematics as a male domain, *J. Soc. Psychology* **141**, 165 (2001).
 - [18] E. A. Gunderson, G. Ramirez, S. C. Levine, and S. L. Beilock, The role of parents and teachers in the development of gender-related math attitudes, *Sex Roles* **66**, 153 (2012).
 - [19] D. I. Miller, K. M. Nolla, A. H. Eagly, and D. H. Uttal, The development of children's gender science stereotypes: A meta analysis of 5 decades of U.S. Draw A Scientist Studies, *Child Development* **89**, 1943 (2018).

-
- [20] K. K. Previs, Gender and race representations of scientists in highlights for children: A content analysis, *Sci. Commun.* **38**, 303 (2016).
- [21] D. I. Miller, A. H. Eagly, and M. C. Linn, Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations, *J. Educ. Psychol.* **107**, 631 (2015).
- [22] K. Finson, Applicability of the DAST-C to the images of scientists drawn by students of different racial groups, *J. Elem. Sci. Educ.* **15**, 15 (2003).