

Blind Spots in Youth DIY Programming: Examining Diversity in Creators, Content, and Comments within the Scratch Online Community

Gabriela T. Richard
Pennsylvania State University
University Park, PA, USA
grichard@psu.edu

Yasmin B. Kafai
University of Pennsylvania
Philadelphia, PA, USA
kafai@upenn.edu

ABSTRACT

Much attention has focused on the lack of diversity in access and participation in digital media available to youth. Far less attention has been paid to the diversity of youth creators and the content that is produced by youth. We examined the diversity of project creators, content, and comments in one of the largest youth programming sites called Scratch (scratch.mit.edu), with over 7 million registered members between ages 6-16, over 10 million posted projects and 16 million comments. We used keyword and webcrawler searches to reveal that only a small number of users (<.01%) self-disclosed their racial and ethnic identities. Case studies further illuminated how project designs and comments delved into race, provided cultural critique or addressed racial harassment. In the discussion, we address these blind spots of diversity in massive online DIY youth communities, discuss methodological limitations, and provide recommendations for future directions in supporting diversity.

Author Keywords

Diversity; Novice Programmers; Collaboration; DIY Production; Media Design; Scratch

ACM Classification Keywords

K.3.1 [Computers and Education]: Computer Uses in Education – *Collaborative learning*; K.3.2 [Computers and Education]: Computer and Information Science Education – *Computer science education*; K.4.1 [Computers and Society]: Public policy issues – *Ethics*; K.8.0 [Personal Computing]: General – *Games*.

INTRODUCTION

Many efforts have concentrated on addressing the underrepresentation of women and minorities in computing activities, online communities, and the technology

workforce. Work has focused on understanding the social and cultural barriers that impede participation [e.g., 5, 37, 38] or on developing new activities such as game design [24] game testing and development [7], electronic textiles [3, 32], and hackathons and coding competitions [43] to recruit women and minorities into computing. Other work has focused on developing programming tools to simplify the mechanics of learning to program and helping novice programmers to become more fluent and expressive with new technologies [3, 31]. Further efforts have focused on developing statewide alliances that bring educators, teachers, and policy makers together [18] to address these disparities. Together these efforts present a formidable collection of interventions to help designers, researchers and teachers understand various issues around diversity and diversifying the pipeline to coding and computing careers. The reason why these findings are important are mirrored in recent discussions about the continued lack of gender and racial diversity in many large technology firms [23].

However, one area that has received far less attention concerns the growing number of online DIY communities where youth, instead of adult professionals, are the project creators and develop digital content. Researchers have long noted that there is a participation gap when it comes to sharing user created content online, and this gap is typically along gender and socioeconomic lines [19]. However, our interest is in whether this trend has continued over time and with the uptick of youth-oriented content creation communities. A recent survey highlighted the growing prominence of such online DIY sites for kids [17] noting that most of the 140 reviewed sites encourage making but not sharing, and thus missing out on elements critical for developing computational participation that engages youth not only in coding but also civic discussions [25]. Scratch [40], a media-rich programming tool and social networking forum [17], engages youth as both creators and commentators. With the increased interest in getting kids into coding, the role that such communities play in providing access, and broadening and deepening participation to computation will only grow. For that reason we need to examine the cultural barriers that might impede their making and contributing not just from the access and participation side but also from the content creator and production side.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CHI'16, May 07-12, 2016, San Jose, CA, USA

© 2016 ACM. ISBN 978-1-4503-3362-7/16/05...\$15.00

DOI: <http://dx.doi.org/10.1145/2858036.2858590>

In this paper, we turn our attention to this neglected aspect of youth online creative production—the racial and cultural diversity (or the lack thereof) of project creators and the project content designed by youth themselves. We chose the Scratch online community because it is one of the most popular youth programming sites and is already widely researched in many aspects of online and offline computational participation [for an overview, see 25]. While gender and geographical demographics of participants and project activities are visible and accessible in part through collected information, public profiles and published statistics based on back-end data analyses, the racial and cultural distribution of users is not collected and available in such public aggregates, and thus invisible to many participants. We contend that such ‘blindspots’ in understanding computational participation need to be revealed and examined so that we can design and support more equitable interventions for youth online creative production. We addressed the following research questions in our analysis: (1) What is the racial and ethnic diversity of project creators on the Scratch website? and (2) How are race and ethnicity addressed in content and comments on the Scratch website? We used keyword and webcrawler searches to examine how youth chose to self-disclose and address their racial and ethnic identities in profiles, projects, and comments. Case studies further illuminated how project designs and comments delved into race, provided cultural critique or addressed racial harassment. In the discussion, we address the blind spots of diversity in youth online DIY communities, discuss methodological limitations, and provide recommendations for future directions.

BACKGROUND

Our interest in the diversity of digital media is not new. It is in fact, a return to an old topic. In the early days of Web 1.0, most energy in this vein was focused on access, getting computers into schools and making networks accessible to all to build the so-called information highway. It didn’t take long to realize that access alone was not enough, and that it also mattered what was available online [34], finding that content was biased in language availability and its audience focus. In the past decade, the focus has shifted from access and content to participation [22], noting that while youth were now able to browse online, most of them were in the less enviable role of content users rather than producers. Attention has now turned to the various ways youth engage in content production [21] ranging from writing fan fiction [1] to making animations [36] and writing code [40]. In particular youth programming communities are growing with the increased interest paid to learning and teaching coding [25].

Most research on representational diversity in this area so far has focused on racial and ethnic representation in commercial online gaming environments where players consume content created by others [13, 39]. While Gee discussed the potential of taking on different “projective” identities online, Nakamura critiqued avatars that

represented non-dominant racial and ethnic groups for often being stereotyped and limiting how much dominant groups could understand diverse experiences. Furthermore, Everett and Watkins [8] proposed that digital games and virtual worlds often create “racialized pedagogical zones” where players learn about race, culture and society through limited stereotypes. Others have further documented the lack of racial diversity and the prevalence of stereotyping in games [48] and virtual worlds [35]. The perils proposed with this lack of representational diversity include a lack of racial awareness, the potential for racist acts in digital and virtual worlds, and non-dominant groups not feeling a sense of identification and belonging. Researchers who examine diverse player experiences have found that when their gender, race, or ethnicity is revealed online – through avatars, biographies or simply using voice chat – women and ethnic minorities are more likely to be harassed or otherwise unsupported [16, 42]. However, not having visible representational diversity online poses similar concerns, as diverse individuals are also less likely to feel that they belong or be seen by others as belonging [35].

Very few studies have examined the racial diversity of creators and content in online sites that are focused on youth DIY production. One early example is a study of racial diversity in content production by teen players in a virtual world called Whyville [26]. The study provided an analysis of player-designed racially diverse body parts for avatars (which were less than 16% at onset), player-written articles in *The Whyville Times* that documented discussions around race, and community responses to the site management’s attempt to create non-racialized avatars to new members. Through a longitudinal analysis, they found that it was difficult to discern the numbers of non-dominant Whyville players, and the availability of diverse body parts increased but was still in the minority, particularly in relation to faces available. While the change toward a racially neutral starting avatar was not met with negativity, some non-dominant users reported experiencing “racist acts” for displaying diverse avatars. These findings overlap with those in commercial gaming that has similarly revealed a lack of representational racial and ethnic diversity, mirroring racial attitudes and stereotyping in society [39]. Many have posited that this lack of consideration for diversity has fed into the “digital divide,” [6] which has changed from lack of technology access, to a “digital identity divide,” or a lack of identification with technology that is often stereotyped around gender and race [15, 42].

The lack of identification with technology is nowhere more visible than in the context of coding communities, as discussed in introduction. Approaches to increase youth participation in programming have focused on providing a ‘low floor’ of entrance for novice programmers, ‘high ceilings’ for making complex projects possible, and ‘wide walls’ by giving room to a range of different interests in projects [41] but also access to like-minded others through

‘open windows’ to share and connect with [4]. A series of recent data mining studies from a random sample of 5,000 users out of 20,000 who logged into Scratch from January to March 2012 examined to what extent these features promote ‘computing for all’ [10, 11, 12]. They found that while 45% of Scratch members posted content on the site, few leveled up to the most extensive forms of social networking and complex programming concepts in their projects. Girls who were equally active in the community [11] were significantly underrepresented amongst the more advanced and experienced coding groups [10]. Particularly relevant to our research is that posting content was a baseline for all visible participation, followed by downloading and only then by commenting. In the Scratch community at least, commenting is one of the first signs of social interaction beyond the more one-sided sharing a project or downloading another’s project [12]. In other words, coding not only encompasses acquisition of technical skills, but also includes engagement with social practices [25].

While these prior investigations in the Scratch community have captured program complexity and its relationship to online participation, they couldn’t examine other equally important factors such as the race and ethnicity of project creators and the actual media content of their programs because this information is neither collected from Scratch members nor contained in backend data. A recent analysis of Scratch projects by Lachney and colleagues [33] found content to be culturally agnostic, heavily leaning toward commercial media from popular video games, television series and toys but less so in terms of culturally relevant content that would appeal to other participant groups. For instance, a simple search of the Scratch archive for the popular video game “Doom” will find hundreds, if not thousands of different programs created and posted by Scratch members while a search for American Indian content will result only in a handful of projects at most, some of them obviously school-initiated as indicated by the creator’s project comments.

These findings are suggestive, as also previous work on online communities revealed, that the Scratch community acts like an affinity space. Affinity groups who coalesce around like-minded interests are powerful learning cultures [13] but they are also exclusive cultures, perhaps not by design but by default. And these exclusions are not just the machination of powerful commercial companies but also replicated by players themselves. It is here where we situate our investigations in the Scratch community: (1) the diversity of creators using registered, self-disclosed, and publicly-available information and (2) the diversity of content and associated discussions using project content and comments. Understanding the presence or lack of “content aware” design in content creation communities is especially important for supporting culturally responsive and expressive participation, as well as fostering diversity within communities at large.

METHODS

Scratch Online Community

Scratch.mit.edu is a massive online community where participants, mostly youth from 11-18 years old, share their computer programs [40]. Kids who share an interest in programming post animations, games, stories, science simulations, and interactive art they have made in the visual programming environment of Scratch. Launched in May 2007 out of the MIT Media Lab, the Scratch site has grown to more than 7 million registered members with over 10,000 Scratch projects uploaded every day [45]. Projects dominate activity and social presence; members often interact with each other through project comments, by remixing projects created by others to make them their own in unique ways, or by curating studios where different projects and project themes are housed and commented on. Member profiles are portfolio-based, showing individuals’ created projects, “favorite” projects, and links to member-created studios (collections of projects) and recent “friends” on their homepage. Traces of the social presence of members take the form of networking residues [17] left on projects and galleries, primarily including comments, love-its and favorites.

Data Collection

Data Sources

To understand diversity within Scratch community, we adopted an iterative approach using different search engines and keywords to identify data on project creators, content, and comments. Our primary data sources were descriptive demographic keywords, followed by case studies of the self-disclosed users and projects. There were three key challenges to our data collection: (1) availability of data; (2) levels of access to public and private data; and (3) levels of public aggregates as reference points. For instance, Scratch members provide information about their age, gender and country but not their race or ethnicity when registering on the site. But only age and country are revealed in public aggregates to all users. Furthermore, Scratch members can create public profiles for their self-selected user names in which they can self-disclose personal data such as age, gender, race and reveal personal interests. Finally, Scratch projects have public notes in which the creators document their project as well as a public comment section in which other Scratch members leave more detailed commentaries on the content and provide critical or technical feedback.

Scratch aggregate data profiles are provided on the site [45] showcasing different statistics on number of project creators, age of project creators, number of projects, number of comments and remixes, as well as geographical distribution. The Scratch community team periodically updates this information. While the team recently released backend data to the public, we chose not to use this data set because it was limited to Scratch 1.0 data (from 2007-2012) and it contained limited information on users (i.e., no data on gender, age or detailed profile information). Further, information about racial identity is not collected from

Scratch members when they register with the site. In addition, even some of the public data is not easily accessible or aggregated and required us to use specialized web crawler searches to become visible.

Most of the information about Scratch project creators and content is publicly available and searchable. The site informs Scratch members that research is conducted on an on-going basis to improve design and functionality, listing names and contacts of the larger research team, and providing links to selected research publications. For this study we accessed publicly available data only and did not identify the users by profile name, to protect identity. For some of the projects, however, it was not possible to anonymize the titles because they were important for displaying the content and issues discussed herein.

Data Collection Period and Phases

The four main phases—feasibility, web crawler search, case studies, and meta analysis—of our data collection occurred from January 2014 through July 2015:

- The *feasibility* phase lasted from January to April 2014. Here we conducted initial searches through the Scratch website to understand how members were self-disclosing racial or ethnic identity. While some Scratch members would discuss their racial or ethnic identity in places other than their profiles (such as in project comments or in blogs), most members self-disclosed this information directly through their projects. At this time, we noted that it was difficult to find certain descriptors (such as “black” or “white” for race) because these terms were used to describe other things in their profiles and resulted in identifying too many unrelated projects. We also found that using the Scratch search engine was limited in its scope in listing projects.
- The *webcrawler* phase lasted from May to October 2014. Here we utilized several webcrawlers that would allow us to go deeper into the site (i.e., listing more results) and also search for particular content more effectively. We identified Screaming Frog SEO Spider webcrawler as the most appropriate software after running several small pilot studies on the keyword data. We iteratively refined our keyword selection based on pilot searches (more detail in next section).
- The *case study* phase lasted from October 2014 to January 2015. Here we cross-referenced the results by looking through each profile and ensuring that the crawler did indeed pick up demographic information from users from the United States. We eliminated profiles that came up as false positives from further analyses because they did not contain demographic information.
- The *meta-content* phase lasted from January 2015 and July 2015. Here we further explored the meta content by looking through the cases to see if the users created projects pertaining to their cultural, racial or ethnic heritage, or related issues (i.e., utilized a “content aware” approach). In many scenarios, users either had no content, or “content

agnostic” content. We focused on cases where there was content related to their sociocultural background.

Data Analysis

We compiled a list of common identifiers, and ran a more comprehensive search with the Screaming Frog SEO Spider webcrawler, which could go more in depth than the internal search engine available on the Scratch website. Our data sources were keywords compiled first through initial searches done through the Scratch website, from which we created case studies, followed by snowball sampling of these keywords through profile descriptors. We looked through profiles and groups to see what data were voluntarily and publicly divulged. For example, we started by exploring keywords such as “African American” or “Latina,” and looked to see what kinds of profiles and projects came up as a result through the Scratch website. From there, we found that members self-identified by other racial and ethnic characteristics, such as “Boricua” (meaning Puerto Rican), or “biracial,” either in users’ profiles or projects. One major reason for using a webcrawler was to limit which sections of the Scratch website we could search. Keywords such as “African American” brought up numerous history projects. For reasons of statistical validity, we limited keyword searches to user profiles. We further limited racial and ethnic descriptors to focus on individuals from the United States, by concentrating on descriptors such as “Asian American,” or adding “United States” to the search term, since international participation was numerous, varied, and outside of the scope of this analysis.

The two most common places where demographics would be located were in member profiles (most common) or on a project. The least common places were in blog posts affiliated with Scratch or in the comments section. After examining 25 case studies using the keywords, we identified that the most reliable demographics data were found in their profiles. After conducting ten subsequent searches, we narrowed down reliable racial/ethnic keywords to the following: “African American,” “Asian American,” “American Indian,” “Boricua,” “byracial” (a misspelling by members), “Chinese American,” “ethnicity,” “half American,” “Hispanic,” “Hispanic American,” “I am/I’m black” (“black” alone revealed thousands of projects involving the color black), “Caucasian,” “I’m white,” “I am white,” “Japanese American,” “Korean American,” “Mexican American,” “mixed,” “part American,” “Filipino United States,” and “racial.” Racial was often used to find descriptors such as “bi-racial” or “multi-racial.” Certain other keywords, such as “Native American” and “Filipino,” had to have results be manually verified after finding that only 25% of the profiles contained the term as a demographic and the majority referenced projects, even after limiting the search to user profiles. The following table 1 provides an overview of the keywords most commonly associated with race and ethnicity.

Race/Ethnicity	Keyword
Multiracial	part American, mixed American, half American, mixed-race, biracial, byracial, racial
Asian American	Asian American, Korean American, Japanese American, Chinese American, Filipino American
African American	African American I am Black I'm Black
Native American	Native American American Indian
Latino / Hispanic American	Hispanic, Hispanic American, Boricua, Mexican American
Caucasian / White	Caucasian, I'm White, I am White

Table 1. Race/ethnicity and keywords associated.

We identified case studies first by looking through the feasibility data, and then as a follow up to our webcrawler search. The main criteria for selecting cases were disclosure of racial/ethnic heritage in members' profile or discussions of their racial/ethnic heritage on the Scratch website. We then analyzed creator profiles, projects, project notes, and comment sections. The three cases included in this paper are convenience samples. The "Latina Princess" Scratch project did not provide any notes while the "Racism Bullying" Scratch project contained very detailed notes where the creator thanked many individuals, but also personalized the project by stating it was based on his life.

The meta-content analysis was conducted by examining projects created by case study members who self-disclosed their demographic information. The search and inclusion criteria here were projects that discussed issues related to cultural, racial or ethnic heritage. Given that it was difficult to search projects under demographic terms (because we were most likely to find school-related history projects), we were limited to snowball sampling within our convenience sample of cases. We analyzed what projects revealed about how Scratch members understand and respond to projects with cultural content.

FINDINGS

What we know about the diversity of project creators, content and conversations of the Scratch online community presents a complex collage of what is visible and what is not visible depending on whether this information is provided by the Scratch members themselves at the time of registration and revealed by Scratch community team, self-disclosed in public profiles, or drawn from backend data by researchers with access privileges.

Diversity of Project Creators within Scratch Community

We start with information about age and location of Scratch members that is self-disclosed by new users at time of registration and publicly collated by the Scratch community team online [45]. In terms of age diversity, the Scratch

community is skewed towards adolescents with the majority of Scratch members between ages 13-15 and then including a long tail of increasingly older users. In terms of geographical diversity, three millions members (43%) of the now seven million registered members are located in the United States, the remainder spread out across the globe [2, 45]. The majority of members in the Scratch community are from the US, UK and Asia and thus could point towards the dominance of English but the translation of Scratch in 40 different languages by national user groups has created a geographically widespread base (see Figure 1). The nationwide promotion and introduction of computer science education in countries such as UK and Asia can, in part, explain the heavy concentration in areas outside of the US.

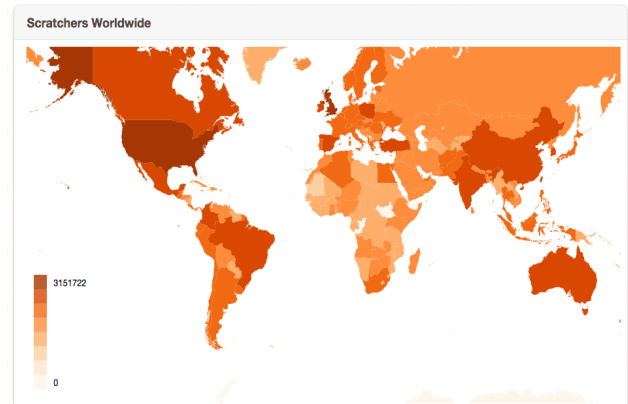


Figure 1. Geographical Distribution of Scratch Community. (Note: Red zones indicate high density)

While information about gender is collected at time of registration and self-disclosed by members in their public profile, the aggregate data is only published in research reports. Initially, 33% of its members were girls [40] but in recent years this percentage has increased to 45% [11]. While Scratch 1.0 only collected information about female or male gender, the registration in Scratch 2.0 now provides an open text field to let new members self-identify their gender. Statistical information about the distribution of gender is not publicly displayed on the site, only available from backend data but Scratch members can choose to reveal their gender in their public profiles. We attribute this recent growth in female membership to popular outreach campaigns such as the Hour of Code [20] that have moved the Scratch membership base beyond the informal users to classrooms which are often more gender-balanced and thus might have introduced a larger number of girls to Scratch and other tools. But we also know from previous research that girls' programming tends to involve less complex concepts [10] thus raising a critical equity issue in participation.

The information about ethnic or racial identity is only available via self-disclosure in public profiles and neither aggregated by Scratch community team, nor collected at the backend and published in research reports. We found that

the vast majority of Scratch members chose to not reveal their racial or ethnic background in their profiles. Based on the keywords used in webcrawler analysis, less than .0001% of the 3,013,098 registered Scratch members in the United States [45] self-disclose their ethnic or racial identity. The majority of these 140 US Scratch members were Latino or Hispanic American (72), followed by Asian American (28), African American (22) Multiracial (13), and Native American (5). Only five Scratch members explicitly self-disclosed themselves as Caucasian/White. A small number of members also included religious background.

A further analysis of these identified 140 Scratch members indeed confirmed that the members had self-disclosed this information in their public profiles. Two different cases illustrate how Scratch members choose to disclose their racial/ethnic characteristics in the “About Me” section which every Scratch member can access by clicking on the username (Figure 2) listed under posted projects on the site.

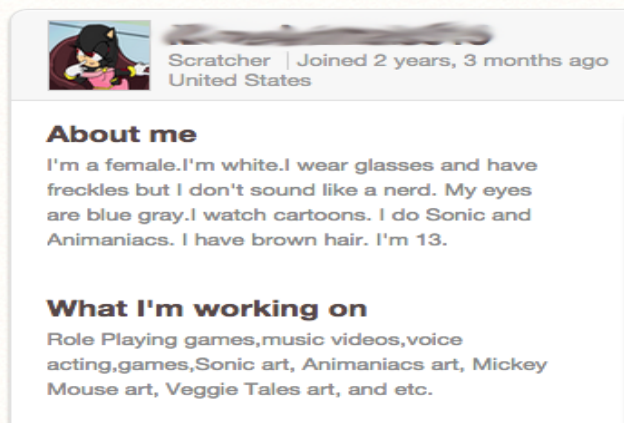


Figure 2. Member self-disclosure in profile.

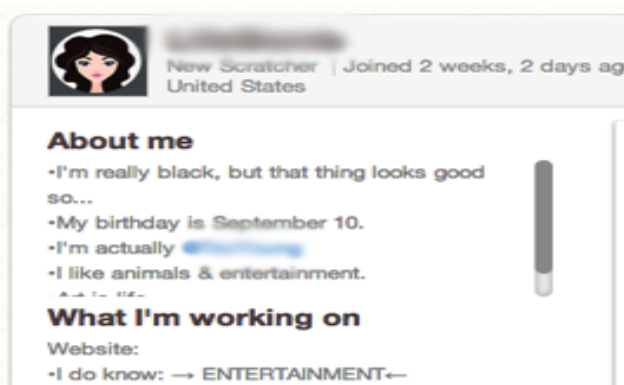


Figure 3. Member self-disclosure in profile.

For instance, Figure 2 showcases a more typical profile on the site where the Scratch member lists age, gender and some physical features including her personal media interests. In contrast, Figure 3 showcases a profile that also lists racial identity but reveals an interesting conflict between the member and the avatar chosen in the her

description. She indicates that she is “really black, but that thing looks good so...” hinting that she chose a white avatar because it looks pleasing to her, but she needs to validate her racial background. This indicates both a concerted action on the part of the user, but also a level of racial pride and need for acknowledgement. We know from prior research that it is not uncommon in this age group to experiment with different identities in their avatar design [26]. The self-created avatars that are part of a Scratch public profile actually provide rich opportunities to create different kinds of graphical character representations.

Diversity in Project Content within Scratch Community

Moving from the project creators to the project content within Scratch community illustrates how the massive number of projects and comments posted both showcases diversity on an easily accessible level *and* leaves out critical issues. According to the latest statistics, over 10 million projects have been posted by the 7 millions members on the site including an even larger number of 16 million comments [45]. These activities are illustrative of the larger social dynamics on the site that foster both content creation and social networking, a rare feature in many kids’ DIY sites [17]. On an average day, looking at the featured projects on the Scratch website, one is likely to find a wide variety of projects in terms of subject areas and themes (see Figure 4). A snapshot of the front page, taken in late 2014, featured mazes, dragons, stories, “Warrior Cats” and other pop culture creations, as well as how-to tutorials created by Scratch members.

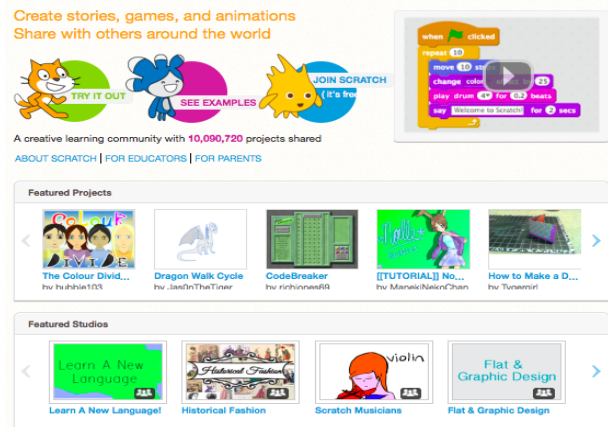


Figure 4. Scratch front page with featured projects.

The one exception on the featured page during data collection (see Figure 4 and Figure 5) is the *Colour Divide* Scratch project, which is actually based on a cartoon role-playing series created by several Scratch members. While the characters are diverse, the *Colour Divide* refers to magical abilities, and not the cultural diversity of the characters themselves. However, it can be seen as an example that takes advantage of providing diverse characters as part of the story and game that stands apart from the majority of featured projects.

This focus of projects based on popular TV shows, games or toys is highly prevalent on the Scratch site reflecting a tendency to gravitate towards violent video game formats and commercial content—indicative of what has been called content agnosticism [33]. While project content on the site also covers academic content such as homework assignments or provides peer assistance for learning how to get up and running with Scratch [12], themes that feature racial, ethnic or cultural diversity are rarely found and thus present a missed educational opportunity for Scratch members to engage with these civic issues.



Figure 5. Project starter screen: *Colour Divide*.

Scratch projects and comments whose content explored race, provided cultural critique or addressed racial harassment were rare finds on the site. Three projects illustrate the different ways in which Scratch members engaged with these issues: (1) the *Latina Princess* project, a cultural critique of how Latinas are portrayed in popular culture; (2) the *Racism Bullying* project, a personal story about racial harassment; and (3) the *Ocealia: Aimi and More!* project, a response to a fantasy series with commentary about the lack of racial diversity in Scratch. These projects showcase a reframing of popular media stories, bringing in personal feelings and understandings around diversity, and engaging the community in issues around diversity. While many comments demonstrate that Scratch members were encouraging, others also displayed negative feedback that reflected a lack of awareness around diversity and related issues.

The *Latina Princess* project (see Figure 6) adopts a fairy tale trope examining whether the brown skinned heroine, named “Maria,” could be a princess. The story starts with a dream sequence where the character is visited by a fairy godmother who wants to take her back to her kingdom. However, Maria quickly responds with confusion stating, “Huh? But I cannot be a princess. I’m Latina.” She discusses various reasons why she thinks so, which have been overwhelmingly influenced by the near invisibility of princesses of color and stereotypes of Latina characters in the media. The project ends with the fairy telling the audience that princesses should be portrayed in a more racially diverse way so that young women of various racial backgrounds can have equal opportunity for the

development of positive and healthy self-esteem. Examining the comments of the two Scratch members who responded, we noticed that both used positive words, such as “cool,” “good points,” and “good luck”—fairly typical comments Scratch members leave for others [12].

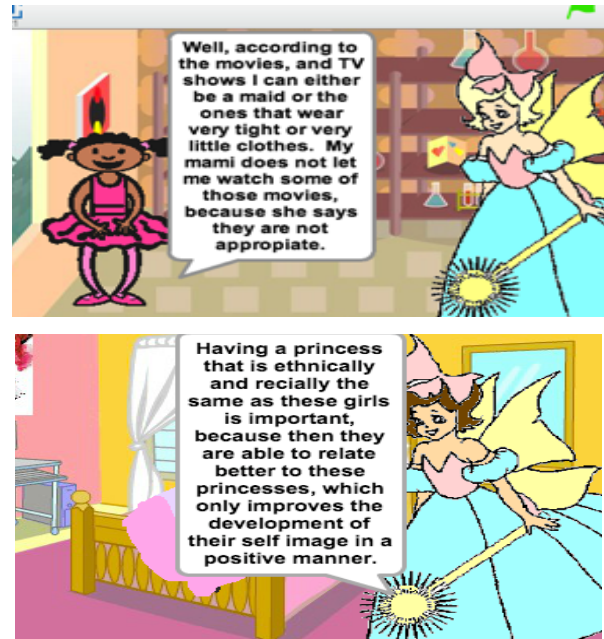


Figure 6. Project screenshots: *Latina Princess*.

But one of the commenters did not understand the final points made by the fairy (whether this is due to resolution issues or not understanding the content is unclear) and the other expressed support but also a sense of inevitable change that would occur because of “census data”:

You raise some good points. People internalize what they see on TV, sometimes not even on the conscious level. Good lick [sic] influencing Disney...I’m sure they are already thinking along those ines [sic]...they pay attention to the Census data!

It is unclear whether the commenter truly believed that there would be more representational diversity with more Latinos in the population of the United States, or if he was expressing sarcasm. However, the comment, in contrast to the project, reflects a lack of awareness of representational issues in media. While the presence of Latinos in mainstream media is on the rise, the point the Scratch member was making was that Latinos are typically found in limited (and often stereotyped) roles that do not include princesses. For example, as of 2015, there has yet to be a Latina princess created by Disney (and this particular Scratch project was created in 2008).

The second project, *Racism Bullying*, was about racial harassment described in project notes as such:

This project says [sic] the story of Racism and ways to prevent it. I have experienced Racism many, many and many times. **ESPECIALLY** [emphasis in original] when I was in ... Elementary school. Do not report this project if it may seem a bit racist to you. Just put it in the comments and I will fix it.

In addition to his project description, of note here as well is his request not to report his project for racism. It is not too much of a stretch to assume that he may have had a problem with his project being reported before. The project itself contained scenes where students of color are ostracized for their diversity, and where Mexican American students, specifically, are targeted for not belonging in the United States. He discusses ways that friends can support and help prevent what he terms “racism bullying.” He even followed up the project with one containing deleted scenes. In the project notes for that project, he stated:

... I tried to combine [the deleted scenes] to a sequel. But since I am leaving scratch forever, I decided to present it as deleted scenes.

Comments were turned off in the deleted scenes project. The original project contained nine positive or supportive comments that included feedback such as “great job” and “very true message,” or suggestions to address grammatical and technical errors. However, there were also four containing negative or disparaging feedback. The negative comments stated things such as “insulting a Mexican isn’t racist” and “I don’t ... believe that there is any need for another sequel...because it’s just giving more examples to be a racist...” Afterward, one of the original creator’s Scratch friends remixed his project.

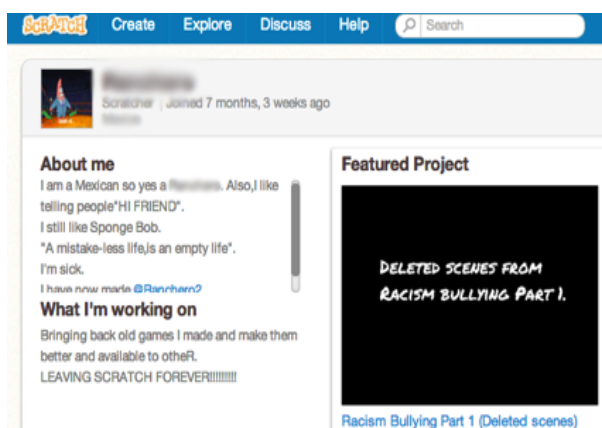


Figure 7. Scratch profile of *Racism Bullying* creator.

Additional comments on this remixed project reflected that one Scratcher thought the original content creator was Black because of the project he made, and, when he is told that he is Mexican, the commenter responds that he feels guilty because calling him Black “is racist.” While it is difficult to determine how the original content creator felt

about these comments based on his responses, which remained positive, he eventually wrote on his profile page that he was “leaving Scratch forever” (see Figure 7), something that was also mentioned in his final “deleted scenes” project. The creator of the *Latina Princess* project similarly never created another project on the original profile.

Finally, the *Ocealia: Aimi and More!* project was created in response to a fantasy series created by another Scratch member based on popular Japanese anime characters (see Figure 8). This project included a new version of a character, called Aimi, who he described as “half Caucasian/half African American but looks a tad bit tanner if that makes sense. She also has light grey hair with emerald shiny green eyes.” While it was impossible for us to discern the racial and ethnic makeup of participating members, the comments around the *Ocealia* project reveal concerns about the lack of racial diversity in Scratch.



Figure 8. Scratch project starter screen with Aimi.

The following excerpted exchange between the Scratch project creator and another Scratch member provides examples of how this was addressed in project comments:

- Scratch User:** yes an african american oc [original character] to join my race thank you
- Scratch Project Creator:** @ScratchUser wut
- Scratch User:** @ScratchProjectCreator Aimi is half african american and so am I. Like actually me.
- Scratch Project Creator:** @ScratchUser 4 of my OCs [original characters] are actually ^^, I'm actually "full" despite not really looking like it because I guess past ancestors?
- Scratch User:** @ScratchProjectCreator omg really
- Scratch Project Creator:** @ScratchUser Yep, yep
- Scratch User:** @ScratchProjectCreator thats really cool
- Scratch Project Creator:** @ScratchUser Really? ^^ Well I guess so :D [smiley face]
- Scratch User:** @ScratchProjectCreator I don't find many african american people on scratch..
- Scratch Project Creator:** @ScratchUser Yeah, but I don't really think about it.
- Scratch User:** @ScratchProjectCreator meh tru
- Scratch Project Creator:** @ScratchUser Oh well

The original commenter (called “Scratch User” to protect identity) was obviously very excited to see a diverse character created, even commenting on the lack of such characters, as well as Scratch members, as far as she is aware. It is then that the project creator informs her that he is also African American, which she finds fascinating because, up until then, she had not met very many diverse Scratch members. However, as soon as she expresses enthusiasm, the creator acknowledges it but dismisses it as unimportant, to which the Scratch user concedes. In this way, this interaction demonstrates both the lack of awareness non-dominant participants have of their presence, and the seeming lack of celebration and acknowledgement within the Scratch community, reflecting a sentiment akin to the “content agnostic” approach [33]. While the creator filled a void by creating a diverse character, her uniqueness as diverse was not seen as important by the creator, along with his and his fellow Scratcher’s racial background, despite its lack of disclosure and prevalence.

DISCUSSION

Our analyses of project creator, content, and comment diversity in the Scratch online community revealed both promising but also problematic findings. On one hand, we see diversity in age (though young adolescents are the largest participating group), in geography with an emphasis on English speaking regions, and increasingly also in the gender ratio of Scratch members. Most of these findings are visibly presented and collected on the site. On the other hand, we found a lack of self-disclosed information about race and ethnicity by projects creators and a dearth of discussions around race, cultural critique or racial harassment. The numbers of these small but significant indicators reveal that at least from a US-focused perspective, and arguably the largest group of contributors, discussions of cultural identity and issues are not part of the Scratch community. What we did not examine here are issues such as religion that might be of prominence in other geographical regions. In the following sections, we discuss further issues of diversity in youth online creative production, methodological limitations in research, and recommendations for designers of online community sites.

Promoting Content Diversity as Community Access

Our focus on project creators and project content is a smaller but much needed slice in examining the larger issues of diversity connected to online communities. In the past much more attention has been focused on the lack of diversity in gender representation. But diversity is important for providing not just access to broadening participation but also for deepening participation [25]. We see multiple benefits in promoting cultural content diversity in youth online creative production sites like Scratch by inviting in more members and enriching the conversations about critical issues. If we want to address the underrepresentation of racial and ethnic minorities in computing fields [37], broadening access and participation

is an important first step. The content on the front portal signals to newcomers and oldtimers alike what topics are deemed valuable and by extension, who should be part of the conversation. If we want to adopt the metaphor of the house with *low floors* for facilitating beginnings, *high ceilings* for allowing complex projects, *wide walls* for addressing multiple interests [41] and *open windows* for increasing social participation [4], then we need to add *broad doors* for allowing in all who are interested. While youth themselves can choose their own content and avatar representation on the Scratch site, it is telling that few of them self-disclose their racial/ethnic identity, design culturally relevant or responsive projects, or even are aware that there is diversity on the Scratch online community.

As to the second point, we see equal benefits in enriching the conversations. If it was difficult for us to identify creators and content connected to race, it is no surprise that other Scratch members might not even be aware of the numbers of Scratch members who may be of the same cultural, racial or ethnic background. With a community like Scratch, there are great opportunities to learn from and about each other and this shouldn’t be just limited to youth from other countries. While many Scratch members were supportive of discussing issues of racial diversity, there was also a strong and vocal unsupportive or unaware stance presented by some. We saw this in the exchange between the two African American members in the *Ocealia: Aimi and More!* project when one minimized the importance of finding other racially diverse members, and in the *Racism Bullying* project, when other Scratch members were quick to challenge the project itself as promoting racism. We also found that, months later, our case study members for the *Latina Princess* project and *Racism Bullying* project, who had also self-identified their ethnicity, either abandoned their profiles or were no longer active on Scratch under that profile.

Obviously, further research is needed to determine why these project creators, and potentially many others, leave the site. Much research has been focused on why Scratch members join the site and choose to stay [2] but next to nothing is known why and how Scratch members choose to leave. We also do not know how racial issues in project content or self-disclosure about racial identity affects the reactions of other Scratch members viewing the exchange. A possible reason why the majority of Scratch projects focus on popular culture may be due to Scratch members wanting an environment of acceptance. This may also be why some choose not to disclose their gender or race publicly. Scholars who study online gaming and associated communities tailored to youth and adult users have found that many choose not to disclose their gender or race online, and moderate their behavior to conform to popularly accepted themes to avoid backlash and harassment [16, 42]. However, hiding diversity is also not a solution, as studies have found the lack of gender and racial diversity can also be a barrier to more diverse participation [35].

Methodological Challenges in Evaluating Diversity

Our search for understanding racial and ethnic diversity on the Scratch site encountered many challenges. While the Scratch community team publishes many statistics about the overall participation and updates them continually, we also observed that certain data about diversity was either not collected (such as race) or not reported (such as gender), at least not publicly on the site. Obviously, the best way to get such representational data about diversity would be at the backend. But this would require asking Scratch members to list their race when they register for an account. What kind of categories would we adopt here in an increasingly multi-racial society? We encounter further problems when operating in a global community that has participants with countries that frame diversity in different ways. While diversity is a global concept, it is situated differently in each country. However, since Scratch community team does not collect this information, a second best approach would be to conduct surveys of the community to see who is using Scratch, as well as what their experiences in the community are. Our analysis is just a first step in understanding what diversity looks like on the Scratch site, but cannot be representational, based on the lack of information.

Lacking access to backend data of Scratch participation or the opportunities to conduct surveys within the Scratch community, we were left to work with search engines. It was difficult to narrow down certain search terms, as well as to conduct searches across large amounts of data sets. But even the use web crawlers had certain limitations as we found out. First, the webcrawler was only as good as the search terms, and even the results had to be manually verified to make sure that they were picking up relevant data. This approach would present significant challenges for a larger data set. Second, the webcrawler would often replicate pages by including the Scratch member's page, and featured project pages, if the featured project was associated with the member's profile. This meant that we had to, again, manually delete if one member had multiple pages. Furthermore, when we wanted to understand the diversity within projects, we often had to analyze it manually. This alone should be indicative enough on how difficult it is for youth to find out how racially and ethnically diverse Scratch members and projects can be.

Recommendations for Increasing Content Diversity

Our findings are not surprising, if we consider the issues around the lack of diversity in technology and computing cultures at large [14, 23, 37, 47]. We also know that providing access to and presence on the site does not guarantee all forms of computational participation [10, 11]. One possible direction is to think about what kinds of supports are offered to foster this access to less represented groups. Research on women gaming communities has shown that supportive communities can help rectify the participation gap [42], as can instilling a sense of ethnic pride for racial and ethnic minority youth in online spaces

[46]. Many emerging initiatives, such as *Girls Who Code* [14] or *Black Girls Code* [47] have further attempted to rectify that lack of women and non-dominant groups in professional coding environments. Further, within Scratch, past efforts have created "collab camps," [44] opportunities for Scratch members to collaborate around different themes and projects. Efforts could be made to focus some of these camps and challenges around diverse content and representational themes.

Other possible directions include focusing our efforts on ways that youth can be more mindful in the content they create. In other words, instead of just taking a content agnostic approach, and policing certain kinds of negative behavior online, we can provide scaffolds and supports that encourage "content aware" design by members within the environment itself. For example, Scratch members are encouraged to write how they have changed an original project when they remix it to make it their own. Similar prompts could be provided to have them think about how their projects engage in discussing culture, civic engagement, or diversity. Another possibility would be to feature positive cultural projects on the main featured projects page. In this way, we can model positive and healthy pathways toward inclusive computing for all with important discussions around the issues that these projects and activities can provide.

CONCLUSION

This study provided a complex collage of what is visible and what is not visible about the racial diversity of Scratch members and content in one of the largest youth online programming and DIY content creation communities. We identified "blind spots" in our understanding of diversity in such communities, as well as new ones that are opened through this exploratory analysis. Moving forward, we need to think about and identify ways that we can diversify technology and computing through communities focused around youth creation and participation. What we know about the racial diversity of project creators, content and conversations of the Scratch online community presents the need for more work in this area, particularly in the areas of backend data collection, surveys of community member participation, and creating opportunities for more critical and civic engagement with youth around diversity and representation.

ACKNOWLEDGMENTS

The data collection, analysis and writing of this paper have been supported by grants from the National Science Foundation (#1027736 and #1450877) to Yasmin Kafai and a Postdoctoral Fellowship to Gabriela Richard by the University of Pennsylvania. The views expressed are those of the authors and do not necessarily represent the views of the National Science Foundation, Pennsylvania State University, or the University of Pennsylvania. Special thanks to Radium Lee Yang for web crawler searches and to Barrie Adleberg for initial data collection.

REFERENCES

1. Rebecca W. Black. 2008. *Adolescents and online fan fiction*. Peter Lang Press.
2. Karen Brennan, Amanda Valverde, Joe Prempeh, Ricarose Roque, and Michelle Chung. 2011. More than code: The significance of social interactions in young people's development as interactive media creators. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications (EdMedia'11)*, 2147-2156.
3. Leah Buechley, Mike Eisenberg, Jaime Catchen, and Ali Crockett. 2008. The LilyPad Arduino: using computational textiles to investigate engagement, aesthetics, and diversity in computer science education. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI'08)*, 423-432.
4. Quinn Burke and Yasmin B. Kafai. 2014. A decade of programming games for learning: From tools to communities. In *The Handbook of Digital Games: Institute of Electrical and Electronics Engineers (IEEE)*, Marios C. Angelides and Harry Agius (eds.). John Wiley & Sons, New York, NY, 689-709.
5. Joanne Cohoon and William Aspray. 2006. *Women and information technology: Research on underrepresentation*. MIT Press.
6. Brady A. Curlew. 2005. Liberal Sims?: Simulated difference and the commodity of social diversity. In *Proceedings of the 2005 Digital Games Research Conference (DiGRA)*, 1-7.
7. Betsy DiSalvo, Mark Guzdial, Charles Meadows, Ken Perry, Tom McKlin, and Amy Bruckman. 2013. Workifying games: successfully engaging African American gamers with computer science. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education (SIGSCE'13)*, 317-322.
8. Anna Everett and S. Craig Watkins. 2008. The power of play: The portrayal and performance of race in video games. In *The ecology of games: Connecting youth, games, and learning*, Katie Salen (ed.). The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning. MIT Press, Cambridge, MA, 141-166.
9. Deborah A. Fields, Michael Giang, and Yasmin B. Kafai. 2013. Understanding Collaborative Practices in the Scratch Online Community: Patterns of Participation among Youth Designers. In *Proceedings of the 10th International Conference on Computer-supported Collaborative Learning (CSCL '13)*, 200-208.
10. Deborah A. Fields, Michael Giang, and Yasmin Kafai. 2014. Programming in the wild: trends in youth computational participation in the online scratch community. In *Proceedings of the 9th workshop in primary and secondary computing education (WiPSCE'14)*, 2-11.
11. Deborah A. Fields, Yasmin Kafai and Michael Giang. 2016. Coding by Choice: A Transitional Analysis of Social Participation Patterns and Programming Contributions in the Online Scratch Community. In *Mass Collaboration and Education*, Ulrike Cress, Johannes J. Moskaliuk & Heisawn Jeong (eds.), Springer, 213-244.
12. Deborah A. Fields, Katarina Pantic, and Yasmin B. Kafai. 2015. "I have a tutorial for this": the language of online peer support in the scratch programming community. In *Proceedings of the 14th International Conference on Interaction Design and Children (IDC'2015)*, 229-238.
13. James Paul Gee. 2014. *What video games have to teach us about learning and literacy*. St. Martin's Press.
14. Amy Golod. 2015. No Boys Allowed: Tackling the Coding Gender Gap. April 2015. US News and World Report, Retrieved August 20, 2015 from: <http://www.usnews.com/news/stem-solutions/articles/2015/04/22/girls-who-code-takes-on-gender-gap-in-computer-science>
15. Joanna Goode. 2010. The digital identity divide: how technology knowledge impacts college students. *New media & society* 12, 3 (May 2010), 497-513.
16. Kishonna L. Gray. 2012. Intersecting oppressions and online communities: examining the experiences of women of color in xbox live. *Information, Communication & Society* 15, 3 (March 2012), 411-428.
17. Sarah M. Grimes and Deborah A. Fields. 2012. *Kids online: A new research agenda for understanding social networking forums*. The Joan Ganz Cooney Center at Sesame Workshop, 1-69.
18. Mark Guzdial, Barbara Ericson, Tom Mcklin, and Shelly Engelman. 2014. Georgia computes! an intervention in a US state, with formal and informal education in a policy context. *ACM Transactions on Computing Education (TOCE)* 14, 2 (June 2014), 1-13.
19. Eszter Hargittai and Gina Walejko. The participation divide: content creation and sharing in the digital age. *Information, Community and Society* 11, 2 (March 2008), 239-256.
20. Hour of Code. Retrieved September 1, 2015 from <https://hourofcode.com/us>
21. Mizuko Ito, Heather Horst, Matteo Bittanti, Danah Boyd, Becky Herr-Stephenson, Patricia G. Lange, C. J. Pascoe, and Laura Robinson. 2008. *Living and Learning with New Media: Summary of Findings from the Digital Youth Project*. John D. and Catherine T. MacArthur Foundation.

22. Henry Jenkins. 2006. *Fans, bloggers, and gamers: Exploring participatory culture*. NYU Press.
23. Stacy Jones and Jaclyn Trop. 2015. See how the big tech companies compare on employee diversity. *Fortune Magazine*. (July 2015). Retrieved September 2, 2015 from: <http://fortune.com/2015/07/30/tech-companies-diveristy/>
24. Yasmin B. Kafai. 1995. *Minds in play: Computer game design as a context for children's learning*. Routledge.
25. Yasmin B. Kafai and Quinn Burke. 2014. *Connected code: Why children need to learn programming*. MIT Press.
26. Yasmin B. Kafai, Melissa S. Cook, and Deborah A. Fields. 2010. "Blacks Deserve Bodies Too!": Design and Discussion About Diversity and Race in a Tween Virtual World. *Games and Culture* 5, 1 (January 2010), 43-63.
27. Yasmin B. Kafai, Deborah A. Fields, and William Q. Burke. 2012. Entering the Clubhouse: Case Studies of Young Programmers. In *End-User Computing, Development, and Software Engineering: New Challenges: New Challenges*, Ashish Dwivedi and Steve Clarke (eds.). Information Science Reference, Hershey, PA, 279-294.
28. Yasmin B. Kafai, Deborah A. Fields, Ricarose Roque, William Q. Burke, and Andres Monroy-Hernandez. 2012. Collaborative agency in youth online and offline creative production in Scratch. *Research and Practice in Technology Enhanced Learning*, 7, 2 (July 2012), 63-87.
29. Yasmin B. Kafai and Kylie A. Pepler. 2011. Youth, Technology, and DIY: Developing Participatory Competencies in Creative Media Production. *Review of Research in Education*, 35, 1 (March 2011), 89-119.
30. Yasmin B. Kafai and Kristin A. Searle. 2011. Safeguarding Play in Virtual Worlds: Designs and Perspectives on Tween Player Participation in Community Management. *International Journal of Learning and Media*, 2, 4 (Fall 2010), 1-14.
31. Caitlin Kelleher and Randy Pausch. 2005. Lowering the barriers to programming: A taxonomy of programming environments and languages for novice programmers. *ACM Computing Surveys (CSUR)* 37, 2 (June 2005), 83-137.
32. Stacey Kuznetsov, Laura C. Trutoiu, Casey Kute, Iris Howley, Eric Paulos, and Dan Siewiorek. 2011. Breaking boundaries: strategies for mentoring through textile computing workshops. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*, 2957-2966.
33. Michael Lachney, William Babbitt, and Ron Eglash. Forthcoming. Software Design in the 'Construction Genre' of Learning Technology: Content Aware Versus Content Agonistic. *Computational Culture: A Journal of Software Studies*.
34. Wendy Lazarus and Laurie Lipper. 2003. *The Search for High-quality Online Content for Low-income and Underserved Communities: Evaluating and Producing What's Needed: an Issue Brief and Action Plan with Research Appendices*. Children's Partnership. Retrieved July 2, 2015 from <http://www.childrenspartnership.org/publications/144>
35. Jong-Eun Roselyn Lee and Sung Gwan Park. 2011. "Whose Second Life Is This?" How Avatar-Based Racial Cues Shape Ethno-Racial Minorities' Perception of Virtual Worlds. *Cyberpsychology, Behavior, and Social Networking* 14, 11 (November 2011), 637-642.
36. Kurt Luther and Amy Bruckman. 2008. Leadership in online creative collaboration. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work (CSCW'08)*, 343-352.
37. Jane Margolis, Rachel Estrella, Johanna Goode, Jennifer J. Holme and Kim Nao. 2008. *Stuck in the shallow end: Race, education, and computing*. MIT press.
38. Jane Margolis and Allan Fisher. 2003. *Unlocking the clubhouse: Women in computing*. MIT Press.
39. Lisa Nakamura. 1995. Race in/for cyberspace: Identity tourism and racial passing on the Internet. *Works and Days* 25/26, 13, 1-2 (1995), 181-193.
40. Mitchel Resnick, John Maloney, Andrés Monroy-Hernández, Natalie Rusk, Evelyn Eastmond, Karen Brennan, Amon Millner Eric Rosenbaum, Jay Silver, Brian Silverman and Yasmin Kafai. 2009. Scratch: programming for all. *Communications of the ACM* 52, 11 (November 2009), 60-67.
41. Mitchel Resnick and Brian Silverman. 2005. Some reflections on designing construction kits for kids. In *Proceedings of the 2005 conference on Interaction design and children (IDC'05)*, 117-122.
42. Gabriela T. Richard. 2014. Supporting visibility and resilience in play: Gender-supportive online gaming communities as a model of identity and confidence building in play and learning. In *Identity and Leadership in Virtual Communities: Establishing Credibility and Influence: Establishing Credibility and Influence*, Donna J. Hickey and Joe Essid (eds.), IGI Global, 170-186.
43. Gabriela T. Richard, Yasmin B. Kafai, Barrie Adleberg and Orkan Telhan. 2015. StitchFest: Diversifying a College Hackathon to Broaden Participation and Perceptions in Computing. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education (SIGSCE'15)*, 114-119.
44. Ricarose Roque, Yasmin Kafai, and Deborah Fields. 2012. From tools to communities: Designs to support

- online creative collaboration in Scratch. In *Proceedings of the 11th International Conference on Interaction Design and Children (IDC'12)*, 220-223.
45. Scratch Statistics. 2015. Retrieved July 1, 2015 from: <https://scratch.mit.edu/statistics/>
46. Brendesha M. Tynes, Adriana J. Umana-Taylor, Chad A. Rose, Johnny Lin, and Carolyn J. Anderson. 2012. Online racial discrimination and the protective function of ethnic identity and self-esteem for African American adolescents. *Developmental psychology*, 48, 2 (March 2012), 343-355.
47. Eric Westervelt. 2015. Disrupting' Tech's Diversity Problem With A Code Camp For Girls Of Color. August 2015. NPR Ed, Retrieved August 30, 2015 from: <http://www.npr.org/sections/ed/2015/08/17/432278262/hacking-tech-s-diversity-problem-black-girls-code>
48. Dmitri Williams, Nicole Martins, Mia Consalvo, and James D. Ivory. 2009. The virtual census: Representations of gender, race and age in video games. *New Media & Society* 11, 5 (August 2009), 815-834.