Curating Computer Science Educational Content with Machine Learning

Analyzing Learner Ratings within an Algorithmic Recommender System

Teresa M. Ober, Mayank Kakodar, Ying Cheng, Philip Gonsalves, Paul R. Brenner, Emmanuel Johnson, Bruno Ribeiro, & Janice Zdankus

University of Notre Dame
Purdue University
YWCA Silicon Valley
Hewlett Packard Enterprise

Machine learning (ML) can in theory be used to personalize educational content by identifying online activities aligned with learners' interests. Yet, are learners' self-reported ratings of activities associated with a machine learning generated recommender score? In the current study we sought to address this question using learner's ratings of activity units (i.e., "badges") based on a conventional 7-item Likert scale administered immediately afterwards within an online app designed to teach computer programming skills. The sample included 78 learners (Mean Age = 13.2 years, SD Age = .84 years, %female = 37.2%) enrolled in schools and after-school programs in the United States. Even after controlling for other factors, such as the position of the recommendation on a list and the number of previous selections the learner had made, as well as certain learner demographics, there was a significant positive association between overall badge approval ratings provided by learners and the recommender score. These findings provide validity evidence in support of the ML-generated recommender score by suggesting badges that the learner is likely to approve of, even when considering other relevant factors that could affect their selection of badges, such as the position of the recommendation on the list and the number of selections made by the learner. Further work should seek to establish whether the likelihood of a learner selecting a highly ranked activity is associated with improvement in more domain general attitudes towards computer programming and whether this association is moderated or robust to certain demographic factors.

Keywords: computer programming, K-12 education, STEM inclusion, recommendation algorithm, machine learning, digital curation

There is an ever-growing demand for a workforce with expertise in computer science and programming. In the United States, careers in computer occupations are expected to grow approximately 11% between 2019-2029, a rate nearly three times higher than the national average of 4% across all job sectors (BLS, 2021). Though there is clearly a demand for a highly skilled workforce, computer science and information technology fields generally do not reflect the general population (Code.org, CSTA, & ECEP Alliance, 2020). As such, there is great untapped potential to attract individuals from groups that are currently underrepresented in computing. Women and African American and Hispanic/Latinx individuals in the U.S. are particularly underrepresented in computing careers (NSF & NCSES, 2019).

The issues of participation in the computing workforce are far reaching. However, educational interventions designed that provide positive experiences in computer science and related fields to young learners have been found to be effective in promoting conceptual understanding (Grover et al., 2015) and motivation (Papastergiou, 2009). Providing educational content in computing that aligns with learners’ background interests may be one effective means by which to trigger learners’ initial interest and initiating a cascade that results in long-term positive orientations towards the subject (Eccles, 2007). With advances in machine learning technology, it is becoming increasingly feasible to match educational content with learners' interests thus providing a fully personalized learning experience (Luan & Tsai, 2021). In the present study, we describe the design of one such educational intervention in the form of an online web application, Curated Pathways to Innovation™ (CPI™). We investigate the associations between middle and high school learners’ ratings of educational content with respect to a score produced by a machine learning algorithm designed to recommend personalized content.

Developing Positive Orientations toward Careers in Computing

Diverse participation in computing careers is therefore not only a societal imperative but may also be necessary for meeting the
demands for a growing work sector. Addressing issues in workforce diversity is complicated and may reflect broader issues affecting participation in the science, technology, engineering, and mathematics (STEM) workforce overall. Despite promising initiatives to promote diversity in the STEM workforce (e.g., Miller & Wai, 2015), participation among individuals from underrepresented groups has not substantially improved over the past decade (Varma, 2018).

Issues facing workforce diversity in STEM, and computing fields, likely begin well before workforce entry. There is growing evidence that disparities in children’s and adolescents’ attitudes towards STEM subjects may contribute to the lack of participation later observed in fields such as computer science and programming (Beyer, 2014; Denner, 2011; Gottlieb, 2018; O’Brien et al., 2015). There is evidence that aspirations for pursuing certain career paths begins to develop as early as elementary school (Tran, 2018) and certainly by middle school (Schuette et al., 2012). There is evidence that attitudes towards pursuing careers in STEM fields such as computing become fixed by grade 8 (Le & Robbins, 2016), though some interventions suggest attitudes towards computing are somewhat malleable in early adulthood (Burnette et al., 2020), as well. Beyond foundational awareness of computing, according to Lent et al. (2008), attitudes that are relevant to choosing careers in such areas as computing include domains of self-efficacy (i.e., perceived abilities in the subject area or relevant tasks), expected outcomes (i.e., perceived likelihood of a personally desirable outcome), interest (i.e., general affinity to pursue the subject even in the face of challenge), and aspirations (i.e., goals and values tied to pursuing the subject). While self-efficacy and outcome expectations may form with even unlimited experience in computing, interest and aspirations tend to evolve over a longer time frame (Lent et al., 2002).

Providing Personalized Educational Content Using Machine Learning

Given that interest in pursuing certain career paths is linked to experience, providing learners opportunities to develop the foundations of career interest is important during the formative years of middle and high school is important, such as computing. As learners’ interest vary widely, providing personalized and curated content is thought to be one effective way to do so. Digital curation refers to the use of tools to select, collect, sort, categorize, and share digital content (Antionio et al., 2012). Past research has found that digital curation leads to learners engaging more productively in digital environments in ways that are thought to increase personal salience, both cognitively and emotionally, in a manner which ultimately leads to better learning outcomes (Tsymbulsky, 2020).

Despite the benefits of providing curated digital content to teach computer programming skills, and the abundance of educational content available to do so, to the best of our knowledge, few prior efforts have been made to synthesize and deliver educational content in a curated and personally salient manner. Systems designed to curate content for educational purposes may be one means by which to maximize the potential of educational content (Raghunath et al., 2018). Particularly given advances in machine learning, this appears to be a promising direction for providing personalized and adaptive computer science education experiences (Luan & Tsai, 2021; Pugliese, 2016).

Curated Pathways to Innovation™

Curated Pathways to Innovation™ (CPT™) is an online platform designed to curate developmentally appropriate and personalized educational content for middle and high school aged learners that teaches computer programming and other STEM skills (see Linnel et al., 2020). By delivering curated culturally responsive content matched to learners’ interests and past ratings of activities, the web app creates personalized learning opportunities for learners who have historically deprived opportunities leading towards careers in computer science.

The CPT™ platform is designed to recommend activities based on learners’ interests and ratings of past activities through a machine learning algorithm. Though the recommender orders both activities and groups of activities in the form of badges, in the present study, we focus specifically on the badges. In particular, the recommender jointly learns matrix-factorization (Jackman, 2009) based representations for learners and activities. A Bayesian model (Lee & Seung, 1999), which utilizes these representations and other features, is then used to model how each learner rated each activity.

During live serving (test time), the recommender is given a learner identifier and a list of potential badges (see Figure 1). The recommender then samples activity ratings from the posterior predictive distribution of the Bayesian model. These values are then mean aggregated to yield the recommendation core for each badge. Since the quantity being modelled is a rating in the closed interval [1.5], accounting for variance, the recommender scores lie in the interval [0-6].

Figure 1. List of recommended badges as it appears to users.

After learners log-in through the web-based portal, they have the option to choose a badge. Learners are then shown their progress in completing activities towards badge completion on a page reflecting a pathway. Achieving a badge indicates that a learner has
research successfully completed certain activities designed to teach a common skill related to computer programming. After completing the activity, they return to CPI™ to rate the activity and take a short comprehension quiz to certify completion. If successful, learners may be shown a gif or short animation related to an area they’ve expressed interest in. Subsequently, they are sent back to the pathway page to continue another activity towards badge achievement. After completing all activities within a badge sequence, the learner receives credit for the badge. Earning badges often came with rewards such as the option to personalize an avatar or the unlocking of new activities. Some badges have prerequisites and thus completing one badge may unlock even more.

Past research involving 610 learners enrolled in middle schools in the United States who used CPI™ throughout one academic year found there was evidence of an immediate improvement in their orientations towards computer programming, particularly with respect to their awareness, self-efficacy, interest, and aspirations for a career in computer programming (Ober et al., under review). Furthermore, the study showed that improvement in attitudes towards computer programming did not differ between female and male learners, nor between learners from underrepresented racial/ethnic groups as their counterparts. Considering these promising initial findings, we examined the effectiveness of the recommendation algorithm based on users’ self-report data.

Research Questions
In the present analysis, we attempt to capture information about the extent to which learners’ attitudes are associated with the recommender score. The research questions (RQs) pursued include the following:

1. **RQ1.** Is the score produced by the recommendation algorithm (i.e., “recommender score”) associated with learners’ ratings of the badge?

2. **RQ2.** If so, is there an association, even when considering other factors including:
   a. number of badges the learner had selected to complete, an estimate of their engagement with CPI™;
   b. position of recommended badge on a list;
   c. information available on the learner about their preferences;
   d. whether or not the learner identified as female or a member of an underrepresented minority (URM) racial/ethnic group in STEM (i.e., Black/African American, Hispanic/Latinx, Alaskan Native or American Indian).

Methods
**Participants.** The sample included 78 learners (MeanAge = 13.2 years, SDAge = .84 years, %female = 37.2%) enrolled in middle and high schools located throughout the state of California in the U.S. Approximately 29.0% of the sample identified a member of category which would be considered an underrepresented minority group in STEM, based on the previously described definition (60.3% White/European American, 27.6% Black/African American, 8.6% Hispanic/Latinx, 3.4% Asian/Asian American). Race/ethnicity data were missing for 20 learners.

**Badges and end-of-badge survey.** As previously noted, throughout the academic year, learners completed activities to achieve badges in CPI™. All learners were expected to complete two initial activities (Digital Awareness, Digital Safety and Citizenship) and after completion of the required activities had options to pursue other activities such as learning block coding in Scratch, foundations of Python or JavaScript, or even basic typing skills. In this sample, the learners completed 15 unique badges, 114 badges completed overall. On average, learners completed 1.46 badges (SD = .85, Median = 1.00). After completing a badge, learners were prompted to complete a 7-item 5-point Likert scale to evaluate their overall approval of the badge activities. As shown in Table 1, item-rest correlations were large, suggesting good internal consistency of the items on the scale.

**Data collection and cleaning.** Data were collected online through the CPI™ platform from partnering schools. Prior to data collection, parental consent and child assent was obtained. Only learners who provided the necessary consent documentation were eligible. To be included in the present analytic sample, learners must have selected and completed ratings for at least one badge and must have provided at least some demographic information in their learner profile (e.g., age, gender, race/ethnicity).

**Analytic procedure.** We conducted a multi-level confirmatory factor analysis to determine that the 7-items used to rate the badge had good internal consistency. Once this was established, we took an average of each rating badge as a continuous scale score indicative of a learner’s overall approval of the badge.

The primary aim of this study was to determine whether the recommender score was associated with learners’ self-reported ratings (RQ1). If so, the results provide some evidence that the recommendation algorithm is sensitive to individual preferences. These results could suggest that the recommendation algorithm was sensitive to learners’ orientations towards computing. Finally, we were also interested in determining whether other factors related to the presentation of badges (i.e., position in list), or learner demographic background (i.e., female and URM status) explained variation in the end-of-badge ratings (RQ2).

Given that the same learners provided multiple ratings and that the same badges were rated by multiple learners, we first conducted a mixed-effects analysis which included a random intercept for learners and badges (Bates et al., 2004). If the random effects were found to be significant, they would then be included in the subsequent analyses. If not, then a more parsimonious model would be chosen, without the fixed effect(s).

Results
After demonstrating reasonably good model fit, an average scale score was derived to reflect the learners’ overall approval of the badge activities, with a more positive score reflecting greater approval. The descriptive statistics for the 7-item badge approval
scale is shown in Table 1. Table 2 shows the descriptive statistics for the recommender score and recommendation position (both unique to each learner-badge pairing), and learner selection (unique to each learner).

As can be seen in both Figure 1 and Table 1, learners generally expressed positive regard for the badges they were recommended. The average scale score across all learner-badge pairings was 3.82, suggesting that learners overall tended to agree with the statements in the 7-item scale. For each of 7 items rating the badges, over half (61% and more) held positive orientations towards the badges.

Table 1. End-of-badge scale descriptives.

| Item wording | M   | SD  | Median | Skew | Kurtosis | Item-rest 
|--------------|-----|-----|--------|------|----------|----------
| 1. The activities I worked on for this badge were interesting. | 3.96 | 1.08 | 4.00 | -0.89 | 3.96 | 0.85 |
| 2. I had fun working on the activities in this badge | 3.95 | 1.08 | 4.00 | -0.90 | 3.95 | 0.81 |
| 3. I am good at the kinds of activities that were in this badge. | 3.81 | 1.14 | 4.00 | -0.74 | 3.81 | 0.75 |
| 4. I learned a lot from the activities I completed for this badge | 3.87 | 1.12 | 4.00 | -0.80 | 3.87 | 0.77 |
| 5. I would like to learn more about the information that was covered in this badge. | 3.62 | 1.19 | 4.00 | -0.55 | 3.62 | 0.84 |
| 6. I paid close attention during the activities for this badge | 3.42 | 1.32 | 3.00 | -0.33 | 3.42 | 0.67 |
| 7. The information I learned in this badge will be useful to me in the future | 3.90 | 1.05 | 4.00 | -0.77 | 3.90 | 0.70 |
| Mean | 3.86 | 0.97 | 4.00 | -0.82 | 3.86 | 0.58 |

Table 2. Badge approval scale descriptives.

<table>
<thead>
<tr>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Skew</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rec. Score</td>
<td>114</td>
<td>3.75</td>
<td>0.52</td>
<td>3.67</td>
<td>2.57</td>
<td>4.99</td>
<td>2.42</td>
</tr>
<tr>
<td>Badge Position</td>
<td>114</td>
<td>1.56</td>
<td>2.20</td>
<td>1.00</td>
<td>1.00</td>
<td>17.00</td>
<td>16.00</td>
</tr>
<tr>
<td>Badge Position (log)</td>
<td>114</td>
<td>0.18</td>
<td>0.54</td>
<td>0.01</td>
<td>0.01</td>
<td>2.83</td>
<td>2.82</td>
</tr>
<tr>
<td>Learner Clicks</td>
<td>114</td>
<td>4.15</td>
<td>3.04</td>
<td>3.00</td>
<td>1.00</td>
<td>16.00</td>
<td>15.00</td>
</tr>
</tbody>
</table>

The descriptive statistics shown in Table 2 suggest several important aspects of the predictors used in the subsequent models. For one, we see that the recommender score is generally bounded between about 2.5 and 5. The descriptive statistics on the badge position suggests that learners tended to select badges with positions closer to the top of the list, where a badge position of 1 indicates that it is the first item on the list. Given the large deviations from normality evident by the large skew and kurtosis values, badge position was log-transformed before it was entered as a predictor in the models. Finally, the number of learners’ “clicks,” which reflects the number of badge selections the learner made overall, suggested that each learner had selected approximately 4-5 badges on average (though did not necessarily provide end-of-badge ratings for all of them).

Table 3. Fit comparisons for models predicting badge approval ratings with random effects.

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>LogLik</th>
<th>Model Comparison</th>
<th>Δdf</th>
<th>ΔAIC</th>
<th>ΔBIC</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A</td>
<td>3</td>
<td>320.79</td>
<td>329.00</td>
<td>-157.40</td>
<td>A v. C</td>
<td>1</td>
<td>1.54</td>
<td>.214</td>
<td></td>
</tr>
<tr>
<td>Model B</td>
<td>3</td>
<td>319.64</td>
<td>327.85</td>
<td>-156.82</td>
<td>B v. C</td>
<td>1</td>
<td>.39</td>
<td>.534</td>
<td></td>
</tr>
<tr>
<td>Model C</td>
<td>4</td>
<td>321.25</td>
<td>332.19</td>
<td>-156.62</td>
<td>C v. C</td>
<td>4</td>
<td>.39</td>
<td>.534</td>
<td></td>
</tr>
</tbody>
</table>

Note. Model A includes a random effect (RE) for “badge”. Model B includes RE for “learner”. Model 0 includes a RE for both badge and learner.

Recommender score is associated with end-of-badge ratings. Tables 3 and 4 show the results of the successive models tested with the overall badge approval score as the outcome. We first compared the models with one random effect (Model A: Badge and Model B: Learner) against a null with both random effects (Model 0). The inclusion of the random effects did not lead to improved fit compared against the nested models. Nevertheless, we wanted to control for any variance they could explain and thus both were included in the subsequent models.

Table 4. Fit comparisons for models predicting badge approval ratings with fixed effects.

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>LogLik</th>
<th>Model Comparison</th>
<th>Δdf</th>
<th>ΔAIC</th>
<th>ΔBIC</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>5</td>
<td>308.96</td>
<td>322.64</td>
<td>-149.48</td>
<td>A v. C</td>
<td>1</td>
<td>14.29</td>
<td>&lt;.001***</td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>6</td>
<td>309.91</td>
<td>326.32</td>
<td>-149.45</td>
<td>2 v. 1</td>
<td>1</td>
<td>1.05</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td>6</td>
<td>310.51</td>
<td>326.92</td>
<td>-149.25</td>
<td>3 v. 1</td>
<td>1</td>
<td>.45</td>
<td>.502</td>
<td></td>
</tr>
<tr>
<td>Model 4</td>
<td>7</td>
<td>243.13</td>
<td>260.23</td>
<td>-114.56</td>
<td>4 v. 3</td>
<td>1</td>
<td>2.20</td>
<td>.15</td>
<td></td>
</tr>
</tbody>
</table>

Note. Models 1-4 do include a RE for both “badge” and “learner”.

Table 5 shows which fixed effects are also included in Models 1-4.

Next, fixed effects were entered with the first being the recommender score unique to each badge per learner (Model 1). As shown in Table 4, this led to an improvement in the model fit, as indicated by a small AIC and small log-likelihood values, as well as a significant χ² test. Given that the recommender score was significant in accounting for variation in the overall badge approval rating (β = .371, t = 3.94, p < .001), we included it as a predictor in subsequent models. The additional predictors were subsequently entered into a regression model as fixed effects in addition to the recommender score.

Table 5. Standardized estimates of model fixed effects for badge approval ratings.

<table>
<thead>
<tr>
<th>Model</th>
<th>Predictor</th>
<th>Standardized Estimate (β)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>Rec. Score</td>
<td>.371</td>
<td>3.94</td>
<td>&lt;.001***</td>
</tr>
<tr>
<td>Model 2</td>
<td>Badge Position (log)</td>
<td>-.097</td>
<td>-1.02</td>
<td>.313</td>
</tr>
<tr>
<td>Model 3</td>
<td>Learner Clicks</td>
<td>.066</td>
<td>2.05</td>
<td>.001***</td>
</tr>
<tr>
<td>Model 4</td>
<td>Status (URM=1)</td>
<td>.096</td>
<td>2.05</td>
<td>.001***</td>
</tr>
<tr>
<td>Model 4</td>
<td>Gender (Female=1)</td>
<td>.022</td>
<td>1.68</td>
<td>.055</td>
</tr>
</tbody>
</table>

Recommendation position is not associated with end-of-badge ratings. Position bias is a rather well recognized issue in research on recommendation algorithms, with items appearing in a more accessible location on a list tending to receive more engagement (Wang et al., 2018). The log-transformed badge position was entered as a fixed effect along with the recommender score (Model 2), which had been found to be significant in the previous model. This did not appear to result in model improvement based on the large AIC, BIC, and comparable log-likelihood values compared to the simpler model with only the fixed effect of badge position (see Table 5). In addition, the χ² test was not statistically significant. Thus, we did not find evidence that the position of the recommended badges affected learners’ end-of-badge scores (β = -.097, t = -1.02, p = .313), when accounting for the recommender score (β = .358, t = 3.97, p < .001). In subsequent models, we did not include the effect of badge position.

Number of learners’ clicks is not associated with end-of-badge ratings. Next, we were interested in determining whether the
number of learners’ badge clicks selections was associated with their overall end-of-badge ratings (Model 3). We anticipated that learners with more selections may be more discriminating in choosing activities and thus have higher ratings. As an alternative explanation, the recommendation algorithm may be better able to classify learners’ recommendations given their relatively lengthier history of past choices.

We did not find evidence to support either explanation given the values of the fit indices relative to Model 1. Inspection of the model coefficients also showed that the effect of the number of learners’ selections was not significant ($\beta = .066, t = .64, p = .525$). Instead, we found that the recommendation algorithm is robust even considering differences in the extent to which learners had used CPI™ ($\beta = .376, t = 3.97, p < .001$).

**Gender and URM status are not associated with end-of-badge ratings.** We included learner demographic information as predictors, along with the recommender score (Model 4). Demographic information included gender (female=1) and URM status (URM=1). Due to the missing data in learners’ race/ethnicity, we were not able to compare this model with the previous models ($N_{\text{learners}} = 58$). Thus, model fit indices are not shown for comparison with the other models in Table 3. Based on the standardized estimates shown in Table 4, we did not find evidence that female ($\beta = -.022, t = -.18, p = .855$) or URM ($\beta = .096, t = .86, p = .394$) learners differed in their end-of-badge ratings when considering the recommender score. However, the recommender score was still significant ($\beta = .397, t = 3.51, p < .001$). This again supports the robustness of association between the recommender score and learners’ end-of-badge ratings, even considering these demographic differences. Aside from the recommender score, no other variables explained significant variation in the outcome.

**Discussion**

In the present study, we found that the recommender scores of badges (comprising units of activities) were significantly and positively associated with learner’s overall approval ratings of the badges. This is even after accounting for the other factors.

**Implications**

There are several implications of the findings from the present study. First, we found some evidence in support of the accuracy of the recommender algorithm in selecting badges a learner is likely to rate highly. That this association is robust despite differences in the position of the badge on the recommendation list, learners’ engagement with the CPI™, and certain demographics further supports the validity of the score in identifying and recommending activities the learner is inclined to enjoy. These findings overall provide immediate implications for the efficacy of the algorithm. There are several long-run implications, as well. While further research is needed, these findings indicate that the recommendation algorithm such as the one used here may provide the benefit of curated content that specifically matches learners’ individualized interests, thus activating a long-term interest development (Eccles, 2007). Doing so may help to tap into the interests of learners who have historically had few opportunities to become interested and remain engaged in computer science education (O’Brien et al., 2015). Ultimately, work in this area may provide a progressive step towards closing gaps in workforce participation and improving diversity within STEM and computing fields. While past literature has discussed the benefit of the curation of educational content for personalizing a learning experience (Raghunath et al., 2018), one novelty of the present study is that it provides a novel application of machine learning to accomplish the curation of content.

**Limitations**

Despite the promising nature of the present findings, there are also several limitations. First, the findings are correlational only. Thus, although we were able to rule out several confounds which appeared to be non-significant, we still cannot be certain whether the badge recommender score is improving learners’ overall experience in using the CPI™. Second, we found that the learners who provided ratings generally tended to rate them higher, as evidenced by the descriptives in Table 1. Thus, learners who may have had a very poor experience may not have provided ratings at all. Third, we note that the sample comprised only 78 learners completing 15 badges. Thus, there is a need to collect additional data from learners, whether or not they would typically opt-in to complete the end-of-badge ratings.

**Conclusion**

Interest in the pursuit of certain career pathways appears to begin before middle school (Schuette et al., 2012; Tran, 2018), and likely even sooner. Given the state of workforce diversity and inclusion in STEM, and computer science (NSF & NCSES, 2019), there is an apparent need to encourage female and underrepresented minority learners to engage in computer science curriculum. Providing curated content recommended by a machine learning algorithm may be one effective and scalable means by which to do so. In the present study, we found that the recommender scores produced by the algorithm were significantly and positively associated with learner’s overall approval ratings of the badges, even when accounting for other factors that could influence this association, such as the those related to the badge (i.e., position), or the learner (i.e., number of badge selections). Thus, we found consistent evidence that the recommendation algorithm appears to be producing a score that is reflective of learners’ ratings of the badges. Future work should track learners’ ratings across a larger sample and extended timeframe to identify any potential long-term changes in distal outcomes, such as learners’ interest and aspirations for a career in computing (Lent et al., 2002), in addition to proximal outcomes such as the ratings of activities and badges.

**References**


