Investigating Visitor Engagement in Interactive Science Museum Exhibits with Multimodal Bayesian Hierarchical Models

Andrew Emerson, Nathan Henderson, Jonathan Rowe, Wookhee Min, Seung Lee, James Minogue, and James Lester

North Carolina State University, Raleigh, NC 27695, USA
{ajemerso, nlhender, jprowe, wmin, sylee, james_minogue, lester}@ncsu.edu

Abstract. Engagement plays a critical role in visitor learning in museums. Devising computational models of visitor engagement shows significant promise for enabling adaptive support to enhance visitors' learning experiences and for providing analytic tools for museum educators. A salient feature of science museums is their capacity to attract diverse visitor populations that range broadly in age, interest, prior knowledge, and socio-cultural background, which can significantly affect how visitors interact with museum exhibits. In this paper, we introduce a Bayesian hierarchical modeling framework for predicting learner engagement with FUTURE WORLDS, a tabletop science exhibit for environmental sustainability. We utilize multi-channel data (e.g., eye tracking, facial expression, posture, interaction logs) captured from visitor interactions with a fully-instrumented version of FUTURE WORLDS to model visitor dwell time with the exhibit in a science museum. We demonstrate that the proposed Bayesian hierarchical modeling approach outperforms competitive baseline techniques. These findings point toward significant opportunities for enriching our understanding of visitor engagement in science museums with multimodal learning analytics.

Keywords: Museum-Based Learning, Visitor Modeling, Multimodal Learning Analytics

1 Introduction

Engagement is a critical component of learning in informal environments such as museums [1–2]. Visitor engagement shapes how learners interact with museum exhibits, navigate the exhibit space, and form attitudes, interests, and understanding of scientific ideas and practices. Recent developments in multimodal learning analytics have significant potential to enhance our understanding of visitor engagement with interactive museum exhibits [3–4]. Multimodal learning analytics techniques can be utilized to create computational models for uncovering patterns in meaningful visitor engagement through the triangulation of multimodal data streams captured by physical hardware sensors (e.g., webcams, eye trackers, motion sensors). Multimodal learning analytics has shown significant promise in laboratory and classroom environments [5–

6], but there has been comparatively little work investigating multimodal learning analytics in informal contexts, such as science museums.

Devising computational models of visitor engagement with interactive science museum exhibits poses significant challenges. Visitor interactions with museum exhibits are brief; dwell times with *highly engaging* exhibits often last only 3–4 minutes [7–9]. Furthermore, museums attract a broad range of visitors of varying age, background, knowledge, and learning objectives. Different types of museum visitors show distinctive patterns of engagement, including how they interact with specific exhibits, as well as how they move about the museum floor [10]. To address these challenges, it is important to utilize computational techniques that make efficient use of available data and account for inherent differences in how visitors engage with interactive exhibits in museums.

In this paper, we present a multimodal learning analytics framework for investigating visitor engagement in science museums that is based upon Bayesian hierarchical models. Bayesian hierarchical models explicitly account for differences in patterns of visitor engagement between separate visitor groups. We focus on visitor interactions with a game-based interactive museum exhibit about environmental sustainability, FUTURE WORLDS. By instrumenting FUTURE WORLDS with multiple hardware sensors, it is possible to capture fine-grained data on visitors' facial expression, eye gaze, posture, and learning interactions to model key components of visitor engagement in science museums. We investigate the relationship between multimodal interactions and visitor engagement by analyzing posterior multimodal parameter distributions of Bayesian hierarchical models that model visitor dwell time with the FUTURE WORLDS interactive exhibit. Results show that Bayesian hierarchical linear models more accurately model visitor dwell time than baseline techniques that do not incorporate hierarchical architectures and yield valuable insights into which features are most predictive for modeling visitor engagement.

2 Related Work

Engagement is a critical mechanism for fostering meaningful learning in museums [7]. Much work on modeling learner engagement has focused on formal educational settings, such as school classrooms [11]. In a museum context, low levels of visitor engagement may appear as shallow interactions with an interactive exhibit, or no interaction at all, whereas high-level engagement can manifest as extended dwell times and productive exploration behaviors. We seek to utilize rich multi-channel data streams to identify patterns of meaningful visitor engagement as defined through visitor dwell time with a game-based interactive exhibit. Dwell time has been used previously to examine visitor engagement with museum exhibits [12–13].

Multimodal learning analytics techniques show significant promise for capturing patterns of visitor engagement in museums. By taking advantage of information across concurrent sensor-based data channels, multimodal learning analytic techniques have been found to yield improved models in terms of accuracy and robustness compared to unimodal techniques [14]. Although these applications have shown significant promise, the preponderance of work on multimodal learning analytics has been conducted in laboratory and classroom settings [15–16]. Using multimodal learning analytics to investigate visitor engagement in informal environments is an important next step for the field.

Traditionally, computational models of learner engagement assume relatively high levels of homogeneity across learners in the training data, which is a natural assumption for classroom settings where all learners are approximately the same age and have similar levels of prior knowledge. However, learners express engagement in different ways depending on a range of factors such as prior knowledge and socio-cultural background, suggesting that group-based differences should be considered when modeling engagement [17]. There are limited examples of research on computational models of engagement that account for these differences. Sawyer et al. used Bayesian hierarchical models to investigate models of learner engagement with a game-based learning environment in both classroom and laboratory settings [18]. We build on this work by adopting a Bayesian hierarchical modeling framework for investigating group-level differences in visitor engagement in a museum context.

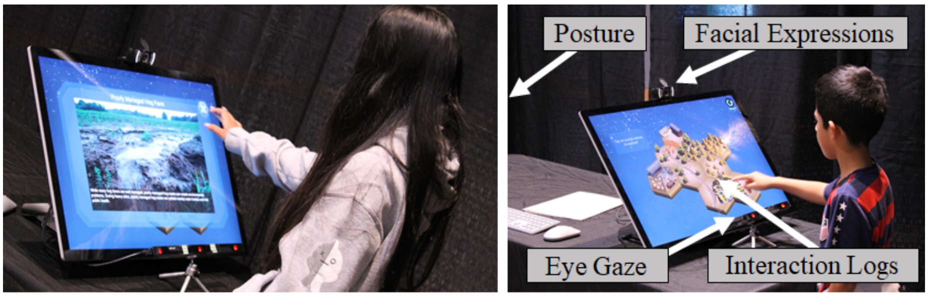

Fig. 1. FUTURE WORLDS museum exhibit capturing multimodal visitor data.

3 FUTURE WORLDS Testbed Exhibit

To conduct data-rich investigations of visitor engagement in science museums, we utilize a game-based museum exhibit called FUTURE WORLDS. Developed with the Unity game engine, FUTURE WORLDS integrates game-based learning technologies into an interactive surface display to enable hands-on explorations of environmental sustainability [19]. With FUTURE WORLDS, visitors solve sustainability problems by investigating the impacts of alternate environmental decisions on a 3D simulated environment (Fig. 1). Learners interact with the environment through tapping and swiping the display to test hypotheses about how different environmental decisions impact the environment's sustainability and future health. Visitors read about different regions of the virtual environment and observe how they are impacted by the learner's actions. The effects of visitors' decisions are realized in real-time within the simulation.

FUTURE WORLDS' focus on environmental sustainability targets three major themes—water, energy (both renewable and non-renewable), and food—and it facilitates exploration of the interrelatedness of these themes. Initial pilot testing with both school and summer-camp groups in a science museum in the southeastern United

States has shown that learner interactions with FUTURE WORLDS enhance sustainability content knowledge and yield promising levels of visitor engagement as indicated by observations of learner behavior [19].

4 Multimodal Data Collection

We leverage a suite of multimodal sensors (e.g., video camera, motion tracking sensor, eye tracker, game logs) to capture visitors' facial expression, body movement, eye gaze, and interaction trace data, respectively, to serve as complementary data sources for inducing computational models of visitor engagement with FUTURE WORLDS. In this work, we focus on modeling visitor dwell time, which is a manifestation of visitors' behavioral engagement, as the ground-truth label of visitor engagement.

4.1 Study Participants and Procedure

We conducted a series of three data collections with museum visitors engaging with the FUTURE WORLDS exhibit at the North Carolina Museum of Natural Sciences in Raleigh, North Carolina. The three groups of visitors were recruited from regional elementary schools from different socio-cultural backgrounds (e.g., race/ethnicity, urban vs. rural, language diversity). Each of the schools served populations where 70% of the students are considered economically disadvantaged. In aggregate, participants included 116 visitors between 10–11 years of age. Each visitor completed a series of questionnaires before and after interacting with FUTURE WORLDS, including a demographics survey, science interest scale, sustainability content knowledge assessment, and engagement survey. Fourteen of the participants did not complete the surveys, which left 47 female and 55 male participants. Approximately 21.6% of the visitors were African American, 8% Asian, 3% Caucasian, 32.3% Latino, and 11.8% American Indian. Visitors interacted with FUTURE WORLDS individually until they were finished or up to a maximum of approximately 10 minutes (*M* = 3.97, *SD* = 2.24). The resulting dataset consisted of complete multimodal data for 86 visitors, following removal of participants with missing data from one or more modalities.

4.2 Multimodal Data Channels

The study utilized a suite of multimodal sensors to gather data on visitor interactions with FUTURE WORLDS. These data streams included facial expression, eye gaze, posture, gesture, and interaction trace logs.

Facial expression. Facial movement data has been widely used to devise computational models for automatically recognizing learning-centered affective states [15]. In our work, we capture facial expression data using video recordings from an externally mounted Logitech C920 USB webcam. The resulting data is analyzed using OpenFace, an open-source facial behavior analysis toolkit that provides automated facial landmark detection and action unit (AU) recognition for 17 distinct AUs [20].

Eye gaze. A growing body of empirical work has demonstrated the importance of eye gaze for modeling learner interactions [21]. To track visitor eye gaze, we utilize a mounted eye-tracking sensor which uses near-infrared light to track eye movements and gaze points during visitor interactions with the interactive exhibit. We automatically identify in-game targets of visitor attention in FUTURE WORLDS using a gaze target-labeling module that processes eye tracking data using ray casting techniques.

Body Movement. Recent years have seen growing interest in research on affective modeling using human body movement data [22–23]. To capture data on visitor posture and gesture, we utilize Microsoft Kinect for Windows v2, a dedicated motion sensing camera that provides skeletal tracking for 26 distinct vertices, in addition to raw pixel data for depth and color camera sensors [24]. The Kinect sensor was mounted on a tripod five feet away from the exhibit and allowed for tracking of body movement.

Interaction Trace Logs. FUTURE WORLDS provides support for detailed logs of learner interactions with the digital interactive exhibit software. The log data consists of timestamped records (at the millisecond level) of visitor taps and multitouch gestures, as well as learning events and simulation states, that arise during visitor experiences.

4.3 Multimodal Features

We extracted several features from each modality to serve as predictors of visitor dwell time. We selected a relatively small number of features for each modality due to the limited size of our dataset. For visitor facial expression, we used AU data captured by OpenFace. We calculated the proportional duration that each AU was exhibited throughout the visitor's interaction with FUTURE WORLDS. Each visitor's facial expression data was standardized and the duration of an AU was recorded if its tracked intensity exceeded one standard deviation above the mean intensity for that AU. Each duration was only recorded if it was present for longer than 0.5 seconds to avoid noise associated with facial micro expressions [25]. We selected 5 AU values: *AU2* (Outer Brow Raiser), *AU7* (Lid Tightener), *AU10* (Upper Lip Raiser), *AU12* (Lip Corner Puller), and *AU14* (Dimpler). These AUs were selected based upon related work on modeling learner engagement with facial expression data [25–27]. We adopted a similar approach to previous work using facial expression for student modeling [25] by scaling the durations of AU data by the total time spent engaging with FUTURE WORLDS.

To capture patterns in visitor attention with FUTURE WORLDS, we used the Tobii EyeX eye tracker to pinpoint areas of interest (AOIs) on the interactive exhibit's display. Visitor fixations on in-game objects exceeding 210 milliseconds in duration were automatically tracked [28]. We aggregated the gaze fixation data to compute the proportion of time visitors spent looking at five categories of in-game objects: virtual locations (AOI-Location), environmental sustainability imagery (AOI-Imagery), environmental sustainability labels (AOI-Labels), environmental sustainability selection menus (AOI-Menu), and user interface elements (AOI-Interface). The *AOI-Location* category included fixations on any of the nine discrete, hexagon-shaped regions of the virtual environment in FUTURE WORLDS. The *AOI-Imagery* category included high-resolution images associated with the exhibit's environmental sustainability content. The *AOI-Labels* category encompassed all textual labels about environmental sustainability topics within FUTURE WORLDS (e.g., text descriptions about renewable vs. non-renewable energy, sustainable farming practices). The *AOI-Menu* category referred to a pop-up menu that appeared when a visitor tapped on a particular location of the virtual environment to learn more about that region or make a change to the region's environmental practices (e.g., add solar panels, introduce organic farming). The *AOI-Interface* category contained user interface elements for navigating the exhibit software (e.g., restart button). Leveraging an approach similar to related work on gaze-enhanced student modeling [29], we calculated the total time spent fixated on each category of in-game element and scaled by the total time spent engaging with FUTURE WORLDS.

To extract features on visitor body movement, we focused on four skeletal vertices tracked by the Microsoft Kinect motion sensor: *Head, SpineShoulder* (upper-back)*, SpineMid* (mid-back)*,* and *Neck*. Selection of these vertices was informed by prior work on multimodal affect detection with motion-tracking sensor data [30]. For each skeletal vertex, we calculated the sum variance of its distance from the Kinect sensor across the visitor's entire interaction with FUTURE WORLDS. Additionally, we utilized the four vertices to calculate the total posture change for each visitor based upon the sum movement of all vertices within the Kinect's coordinate tracking space.

For interaction log features, we calculated the total number of times the visitor tapped on the FUTURE WORLDS exhibit's touch display (*Total Taps*) and the total number of times the visitor tapped to examine environmental sustainability imagery and labels (*Total Info Taps*). The two interaction log features were computed by scaling the above measures by the total dwell time for that visitor (i.e., taps per second), which measured how actively participants interacted with FUTURE WORLDS and its embedded environmental sustainability content.

In sum, we extracted five facial expression features, five eye gaze features, five body movement features, and two interaction log features for a total of 17 multimodal features for this analysis.

5 Bayesian Linear Models

To predict visitor dwell time with the FUTURE WORLDS exhibit, we induced linear models using Bayesian Lasso regression. Lasso regression is a regression analysis method that privileges simpler models by forcing a subset of model coefficients to be set to zero, which serves as a form of feature selection and regularization [31]. We utilized a Bayesian framework to incorporate prior distributions for parameter estimation, account for uncertainty in modeling, and share information across groups of data. Because our dataset contained multimodal data from 86 participants, linear models provided a natural machine learning framework to prevent overfitting and support parameter interpretability. We implemented Bayesian linear models using double exponential prior distributions on all feature coefficients, serving as a form of L1 (Lasso) regularization to limit the number of features utilized in the induced models.

In addition to utilizing prior distributions for model parameters, we also used a logarithmic link function in the regression model to better predict visitor dwell time. In standard Bayesian linear regression, a normal distribution is used to model the relationship between the predictor variables and the dependent variable. The mean of this distribution is the linear combination of the input features and their coefficients. Due to use of the normal distribution, the predictions can be negative. In our case, dwell time cannot be a negative value, so we exponentiate the linear combination of features and coefficients before using it as the mean of the normal distribution. Varying the link function is a form of generalized linear modeling [31]. The formulation for the base linear regression used in our analysis is as follows:

$$Y_i \sim Normal(\mu_i, \sigma^2), \text{ where } log(\mu_i) = \alpha + \textstyle\sum_{k=1}^{p} X_{ik}\beta_k \qquad (1)$$

$Y_i$is the dwell time for visitor i. α is a fixed intercept added to all predictions in the regression, X_{ik} is the value of the input feature k for student i, β_k is the coefficient for feature k, p is the total number of features (of which there are 17), and $\sigma^2$is the fixed variance used for all predictions.

5.1 Baseline Models

We investigated two baseline models using the regression formula (Equation 1) described above for modeling visitor dwell time. First, we use a *Pooled Model*, where all visitor data was grouped together and treated equally. Second, we used a *Group-Specific model*, where a separate linear model was trained on each visitor group. The Pooled Model loses information about the individual groups and does not characterize group-based differences in visitor interest, background, or demographics. This can lead to underfitting of the data. The Group-Specific model is a more specialized form of the regression model, where each visitor group has its own distinct set of model parameters. In comparison to the Pooled Model, this approach risks overfitting the data and is unlikely to generalize effectively due to the limited number of data samples per group and inherent differences between the visitor groups.

5.2 Bayesian Hierarchical Model

The regression formula (Equation 1) assumes that the residual variance for all visitor observations are the same. In many contexts this is a reasonable assumption, but in a museum setting, different groups of visitors may arrive with highly different socio-cultural backgrounds, interests, knowledge levels, and learning objectives, among other relevant characteristics. Different groups of visitors may not only spend different amounts of time at exhibits, but their dwell times may have higher or lower variance depending on the group. Thus, it is important that the multimodal models of visitor engagement account for these differences, and therefore treat the error variances differently in the regression formulation. The assumption of equal variance by standard linear models, or *homoskedasticity*, can result in reduced model fit and information loss when the observations come from groups. We propose an extension to Equation 1 to incorporate a learned variance parameter that is unique to each visitor group to ensure that the variance of the residual errors is treated differently depending on the group from which the visitor came. To avoid overfitting to the visitor groups, we used a shared latent distribution to model the three groups' variance parameters. This *Bayesian hierarchical model* is shown below:

$$Y_i \sim Normal(\mu_i, \sigma_g^2), \text{ where } log(\mu_i) = \alpha + \textstyle\sum_{k=1}^{p} X_{ik}\beta_k \qquad (2)$$

The only difference in this regression formulation compared to Equation 1 is that the variance, σ_g^2, varies based on the school group, g.

6 Results

The predictive models of dwell time were trained and compared using student-level leave-one-out cross-validation. We used cross-validation to compare the performance of the Pooled Model, the Group-Specific Model, and the Hierarchical Model. We report R^2, root mean squared error (RMSE), and mean absolute error (MAE) averaged across each cross-validation fold. The performance of each model is reported on the entire dataset as well as the performance for each visitor group.

Each model was trained using Markov chain Monte Carlo (MCMC) sampling in R using the JAGS framework [32]. To check the convergence of the sampling, we used the Gelman-Rubin diagnostic, which is commonly used for evaluating MCMC convergence [33]. For each of the models, we drew 3,000 MCMC samples after omitting the first 1,000 for burn-in. The process of burn-in is performed to ensure the convergence of the Markov chain in MCMC sampling. The final predictive models used the means of the 3,000 samples for each model parameter. Within each of the predictive models, the coefficients of the features, βs, are assigned a prior distribution. For each β, we used a double exponential prior with mean 0 to operate in the same manner as Lasso regression priors. This encouraged many of the feature coefficients to be as close to 0 as possible, resulting in only a few selected features as significant. The group-level variance parameters, σ_g^2, also used a shared prior distribution to relate information across groups. We chose the Gamma distribution with shape and scale parameters equal to 0.1. Each of the prior distributions chosen for this work were relatively uninformative and thus weak. This forced the posterior distributions of the model parameters to be largely affected by the data rather than our prior beliefs.

6.1 Predictive Accuracy

We compared the accuracy of the three Bayesian linear models: the Pooled Model, Group-Specific Model, and Hierarchical Model. Table 1 shows the results for each model in predicting visitor dwell time (seconds). The Hierarchical Model outperformed both the Pooled and Group-Specific models for all visitor groups. For Group 1, the Pooled Model outperformed the competing models, but for Groups 2 and 3, the Hierarchical Model performed best with respect to the three evaluation metrics.

The Group-Specific models were each trained on data from a single group, and then each model was evaluated only using data from that group. The total predictive performance of the Group-Specific Models was calculated by aggregating the predictions of each of the three models and calculating R^2, RMSE, and MAE with the total data. An explanation for why this modeling approach performed relatively poorly its risk of overfitting to a specific group; each visitor group only consisted of 20–40 visitors. Pooling the data and ignoring group-level characteristics yield good results but risks underfitting the data by losing group-specific information about the visitors. The Hierarchical Model takes advantage of both modeling approaches by incorporating group-level information but keeping all data instances pooled using a shared prior for the group-level variance. An alternative approach to hierarchical modeling is to train a set of feature coefficients for each visitor group. However, this approach would multiply the number of model parameters by the number of visitor groups, which risks poor performance due to the limited size of the data sample.

Table 1. Predictive performance of the three linear models.

Model Type	Context	R^2	RMSE	MAE
Pooled	*All Groups*	0.514	93.720	68.334
	Group 1	**0.425**	**85.846**	**72.532**
	Group 2	0.727	70.429	47.583
	Group 3	0.370	118.319	82.637
Group-Specific	*All Groups*	0.285	110.882	81.457
	Group 1	0.303	96.270	74.216
	Group 2	0.685	75.616	54.567
	Group 3	-0.116	157.409	117.060
Hierarchical	*All Groups*	**0.536**	**91.593**	**67.690**
	Group 1	0.411	88.488	72.649
	Group 2	**0.742**	**68.444**	**47.338**
	Group 3	**0.428**	**112.713**	**80.582**

6.2 Posterior Distributions of Model Parameters

Bayesian models allow summarization and comparison of model parameters by using the MCMC samples that were directly taken from their posterior distribution. As the Hierarchical Model outperformed both the Pooled and Group-Specific models, we summarize the model parameters' posterior distributions of the Hierarchical Model.

Table 2. Posterior parameter distributions for Bayesian Hierarchical linear model.

	Mean	SD
Intercept	5.344	0.046
AU12	-0.197	0.050
AOI-Interface	-0.197	0.080
Total Position Change	-0.151	0.052
AU7	-0.130	0.049
Head Variance	0.082	0.095
AOI-Labels	0.081	0.040
AU2	-0.080	0.044
Total Info Taps	0.068	0.056
Total Taps	-0.060	0.043

Table 2 displays the mean and standard deviation (SD) for each of the model parameters from the Hierarchical Model. Since each model induced double exponential priors on the feature coefficients, many of the features resulted in non-significant coefficients. We report the 10 features with the largest coefficients in terms of absolute value, including the model intercept, noting that features from each modality were chosen as being significant. The remaining features had posterior distributions that resulted in a mean of 0. The significant features for the posture modality were *Total Position Change* and *Head Variance*. For eye gaze, the significant features were *AOI-Labels* and *AOI-Interface*. For facial expression, the features were *AU12*, *AU7*, and *AU2*. The features for the interaction log modality were *Total Taps* and *Total Info Taps*.

7 Conclusion and Future Work

Multimodal learning analytics offers significant potential to advance our understanding of museum visitor engagement. However, museums pose distinctive challenges for modeling learner engagement, including the brief duration of visitor dwell times, as well as visitor populations that range broadly in age, prior knowledge, and socio-cultural background. To address these challenges, we have introduced a multimodal Bayesian hierarchical modeling framework for modeling visitor engagement with interactive science museum exhibits. Leveraging multimodal data on visitor interactions with an interactive game-based exhibit for environmental sustainability education across three diverse groups of visitors, we found that Bayesian hierarchical models outperform competing baseline methods. Furthermore, results indicate that features from each modality contributed significantly toward predicting visitor dwell time, underscoring the promise of multimodal learning analytic techniques for modeling visitor engagement.

There are several promising directions for future research. First, extending multimodal models of visitor engagement beyond predicting visitor dwell time to capture patterns of visitors' cognitive, affective, and behavioral engagement is a key next step. Furthermore, adapting multimodal learning analytic techniques to account for the "messiness" of free-choice learning, including fluid grouping at exhibits [12] and complex patterns of movement across the museum floor [10], is an important challenge. Extending this work to other science museums as well as other informal learning contexts (e.g., science centers, aquariums, zoos, and other public spaces) will help reveal and strengthen the generalizability of this approach. Finally, it will be critical to investigate how multimodal learning analytics can inform iterative cycles of design and development by exhibit designers, as well as best practices of museum educators to enhance high-quality visitor engagement in science museums.

Acknowledgements.
The authors would like to thank the staff and visitors of the North Carolina Museum of Natural Sciences. This research was supported by the National Science Foundation under Grant DRL-1713545. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Hein, G.: Learning science in informal environments: People, places, and pursuits. Museums and Social Issues 4(1), 113-124 (2009).
2. Falk, J., Dierking, L.: Learning from museums. Rowman & Littlefield (2018).
3. Blikstein, P., Worsley, M.: Multimodal learning analytics and education data mining: using computational technologies to measure complex learning tasks. Journal of Learning Analytics 3(2), 220-238 (2016).
4. Oviatt, S., Grafsgaard, J., Chen, L., Ochoa, X.: Multimodal learning analytics: Assessing learners' mental state during the process of learning. The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition, 2, 331-374 (2018).
5. Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L., Zhao, W.: Detecting student emotions in computer-enabled classrooms. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence, pp. 4125–4129 (2016).
6. DeFalco, J., Rowe, J., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B., Baker, R., Lester, J.: Detecting and addressing frustration in a serious game for military training. International Journal of Artificial Intelligence in Education 28(2), 152–193 (2018).
7. Diamond, J., Horn, M., Uttal, D.: Practical evaluation guide: Tools for museums and other informal educational settings. Rowman & Littlefield (2016).
8. Lane, H., Noren, D., Auerbach, D., Birch, M., Swartout, W.: Intelligent tutoring goes to the museum in the big city: A pedagogical agent for informal science education. In: International Conference on Artificial Intelligence in Education, pp. 155-162. Springer, Berlin, Heidelberg (2011).
9. Long, D., McKlin, T., Weisling, A., Martin, W., Guthrie, H., Magerko, B.: Trajectories of physical engagement and expression in a co-creative museum installation. In: Proceedings of the 12th Annual ACM Conference on Creativity and Cognition, pp. 246-257 (2019).
10. Shapiro, B., Hall, R., Owens, D.: Developing and using interaction geography in a museum. International Journal of Computer-Supported Collaborative Learning 12(4), 377-399 (2017).
11. Halverson, L., Graham, C.: Learner engagement in blended learning environments: A conceptual framework. Online Learning 23(2), 145-178 (2019).
12. Block, F., Hammerman, J., Horn, M., Spiegel, A., Christiansen, J., Phillips, B., Diamond, J., Evans, E., Shen, C.: Fluid grouping: Quantifying group engagement around interactive tabletop exhibits in the wild. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 867-876 (2015).
13. Knutson, K., Lyon, M., Crowley, K., Giarratani, L.: Flexible interventions to increase family engagement at natural history museum dioramas. Curator: The Museum Journal 59(4), 339-352 (2016).
14. Baltrušaitis, T., Ahuja, C., Morency, L.: Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence 41(2), 423-443 (2018).
15. Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L., Zhao, W.: Detecting student emotions in computer-enabled classrooms. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence, pp. 4125–4129 (2016).
16. Aslan, S., Alyuz, N., Tanriover, C., Mete, S., Okur, E., D'Mello, S., Arslan Esme, A.: Investigating the impact of a real-time, multimodal student engagement analytics technology in authentic classrooms. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1-12 (2019).
17. Archambault, I., Dupéré, V.: Joint trajectories of behavioral, affective, and cognitive engagement in elementary school. The Journal of Educational Research 110(2), 188-198

(2017).

18. Sawyer, R., Rowe, J., Azevedo, R., Lester, J. Modeling player engagement with Bayesian hierarchical models. In: Proceedings of the 14th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, pp. 215–221 (2018).
19. Rowe, J. P., Lobene, E. V., Mott, B. W., Lester, J. C.: Play in the museum: Design and development of a game-based learning exhibit for informal science education. International Journal of Gaming and Computer-Mediated Simulations 9(3), 96-113 (2017).
20. Baltrušaitis, T., Robinson, P., Morency, L.: OpenFace: an open source facial behavior analysis toolkit. In: Proceedings of the 2016 IEEE winter conference on applications of computer vision, pp. 1–10. IEEE (2016).
21. Aung, A., Ramakrishnan, A., Whitehill, J.: Who are they looking at? Automatic eye gaze following for classroom observation video analysis, pp. 166-170 (2018).
22. Henderson, N., Rowe, J., Mott, B., Brawner, K., Baker, R., Lester, J.: 4D affect detection: Improving frustration detection in game-based learning with posture-based temporal data fusion. In: Proceedings of the 20th International Conference on Artificial Intelligence in Education, pp. 144-156, Chicago, Illinois (2019).
23. Patwardhan, A., Knapp, G.: Multimodal affect recognition using kinect. arXiv preprint arXiv:1607.02652 (2016).
24. Zhang, Z.: Microsoft kinect sensor and its effect. IEEE multimedia, 19(2), 4–10 (2012).
25. Sawyer, R., Smith, A., Rowe, J., Azevedo, R., Lester, J.: Enhancing student models in game-based learning with facial expression recognition. In: Proceedings of the 25th conference on User Modeling, Adaptation and Personalization, pp. 192-201 (2017).
26. Grafsgaard, J., Wiggins, J., Vail, A., Boyer, K., Wiebe, E., Lester, J.: The additive value of multimodal features for predicting engagement, frustration, and learning during tutoring. In: Proceedings of the 16th International Conference on Multimodal Interaction, pp. 42-49 (2014).
27. Vail, A., Wiggins, J., Grafsgaard, J., Boyer, K., Wiebe, E., Lester, J.: The affective impact of tutor questions: Predicting frustration and engagement. In: International Educational Data Mining Society (2016).
28. Rayner, K., Li, X., Williams, C., Cave, K., Well, A.: Eye movements during information processing tasks: Individual differences and cultural effects. Vision research 47(21), 2714-2726 (2007).
29. Emerson, A., Sawyer, R., Azevedo, R., Lester, J.: Gaze-enhanced student modeling for game-based learning. In: Proceedings of the 26th ACM Conference on User Modeling, Adaptation and Personalization, pp. 63–72, Singapore (2018).
30. Grafsgaard, J., Boyer, K., Wiebe, E., Lester, J.: Analyzing posture and affect in task-oriented tutoring. In: Proceedings of the 25th Florida Artificial Intelligence Research Society Conference, pp. 438–443 (2012).
31. Reich, B., Ghosh, S.: Bayesian statistical methods. CRC Press (2019).
32. Plummer, M.: JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd International Conference on Distributed Statistical Computing, pp. 1–10 (2003).
33. Gelman, A., Rubin, D.: Inference from iterative simulation using multiple sequences. Statistical Science. 7, 457–511 (1992).