Towards Assessing Visitor Engagement in Science Centres and Museums

Wolfgang Leister, Ingvar Tjøstheim, Trenton Schulz Norsk Regnesentral

Oslo, Norway

{wolfgang.leister, ingvar.tjostheim, trenton.schulz}@nr.no

Göran Joryd Expology Oslo, Norway goran@expology.no

Abstract—Currently, it is difficult to assess the engagement of visitors in science centres and museums for specific installations. We intend to measure how well individual installations work by using non-intrusive assessment technologies. This paper lays out the assessment framework for this goal. The article presents the Visitor Engagement Installation profile that characterises installations along six dimensions. An assessment framework that consists of four layers is presented and explained. First findings of the assessment of a selected installation are presented.

Keywords—assessment; installations; science centres; museums; visitor engagement.

I. INTRODUCTION

Science centres and museums present exhibitions, installations, and educational programmes that are supposed to engage visitors for self-education on a subject and to inspire the visitors to learn more. However, there is little data showing how well these installations perform regarding their goal to transfer knowledge to the visitors other than the use of longitudinal studies [1]. Similarly, there is little data to determine whether adjustments of installations have the wanted effect on a visitor's engagement.

The main objective of our work is to measure the performance of installations, but we cannot, in general, measure this directly. Instead, we assess the experience of visitors and groups of visitors while they use the installation and retrieve parameters and objective data from the installation and its context. We also want to avoid time-consuming observations by the museum staff and keep intrusive methodologies, such as questionnaires, to a minimum.

In our research, we argue that we can assess dimensions of engagement towards an installation by means of subjective assessment and automated observations of technical data from the installations, physiological data of the visitor, camera data, behaviour, etc. These data are used to estimate the performance of the installation, and whether adjustments of such installations contribute to a better engagement and experience.

First, we present an overview of related work, showing both the installation-centric and visitor-centric view of studies (Section II). Then, we show the approach of our proposed framework for assessing engagement (Section III). We present the Visitor Engagement Installation (VEI) profile to characterise installations using six dimensions (Section IV). An assessment of a selected installation follows (Section V). Finally, we present our conclusion (Section VI).

II. RELATED WORK

Science centres are informal learning environments [2] that are distinct from classrooms because they offer free-choice learning [3][4], i.e., visitors can choose which activities to participate in and they can leave at any time.

Lindauer [5] presents a historical perspective of methodologies and philosophies of exhibit evaluations. Lindauer mentions only a few methods that perform measurements using simple metrics of counting or measuring time. In the literature, the majority of evaluations in science centres deals with the assessment of learning, often using a longitudinal approach, i.e., observing a subject or installation over time. Šuldová and Cimler [6] suggest that engagement can be assessed more instantaneously and be used as a part of learning assessment, supporting Sanford's [7] claim that "some compelling evidence links visitor engagement to learning".

We align the literature along two axes, as illustrated in Figure 1: the vertical axis denotes the span between longitudinal and instantaneous assessment; the horizontal axis denotes whether the assessment is *visitor* or *installation*centric. In general, assessing an installation also needs to take an assessment of the visitor into account.

A. Visitor-Centric View

Dierking and Falk [8] present the Interactive Experience Model, which is a visitor-centric model. They define the interactive experience influenced by three contexts: I) the personal context, 2) the physical context, and 3) the social context. Falk and Storksdieck [9] use the principle of identity-related motivation that places visitors into five identity types: I) the explorer; 2) the facilitator; 3) the professional and hobbyist; 4) the experience seeker; and 5) the spiritual pilgrim. This line of visitor studies has been extensively studied [10][11].

Barriault and Pearson [12] present frameworks that analyse the learning experience more instantaneously by identifying learning-specific behaviour observed by cameras and microphones installed within an installation. Šuldová and Cimler [6] refine these methods, but still depend on manual analysis.

B. Installation-Centric View

In the installation-centric view, the science centre assesses installations rather than the visitors. The developers of installations need to consider the aspects of attractiveness, usability, being educational, etc. Young [13] suggests that developers need to be an advocate for the visitors and think as a visitor and recommends a cyclical development process. Allen [14]



Figure 1: Classification of selected work in visitor studies

presents a study of three different versions of an exhibit for the purpose of studying dimensions of interactivity.

In longitudinal visitor studies, observations and sensemaking [15] are often used. In sense-making, qualitative mental models, understanding events, and interpretation of situations in an iterative process (e.g., the data/frame theory of sensemaking [16][17]) are in the foreground whereas we are interested in concrete measurements and quantitative and descriptive data based on machine-retrievable data and questionnaires that allow us to get an instant result.

III. APPROACH

Installations in museums and science centres are complex systems that need to perform in their context together with the visitors. We take an installation-centric approach over a visitor-centric approach since we are interested in how the installations and potential changes of installations will perform. Also in the installation-centric view, it is important to observe visitors, study what they do, and determine whether the installations work as intended.

To assess the engagement for an installation, we developed an assessment framework that takes various types of data into account. While we are creating an estimation model, we need all available data. After we've created a suitable estimation model, our intention is to abstain from intrusive data collection as much as possible. We use a machine learning approach [18] to establish the model.

Developers and owners of installations are interested in what to change once an installation is assessed. This can be achieved by using the VEI profile presented in Section IV. The idea is to characterise an installation along six dimensions that one can adjust. Whether such adjustments are successful can be evaluated in a new assessment.

A. Assessment Framework

We propose an assessment framework that uses objective assessment, physiological responses, and estimation models to derive evidence of how a visit is perceived for individuals and groups of subjects.

An important requirement is that the assessment methods are not perceived as being intrusive. Intrusive assessment methods are usually only applicable in a lab setting, as they reduce the quality of experience (QoE) and, thus, impact the result of an assessment negatively.

Engagement and visitor experience cannot be measured directly. They are latent constructs. From measurable data and an estimation model trained by our machine learning approach we intend to derive a measure of experience of the visitors using an installation. It is similar to a satisfaction index and can be used to evaluate an installation.

B. The Layers of the Assessment Framework

Our assessment framework (Figure 2) consists of four layers: Layer I: the Scenario Layer presents the artefact, the subject, the action or interaction of the subject, other subjects, and, to some extent, observers; Layer II: the Data Collection and Observer Layer describes which data are collected from the elements of the scenario. Layer III: the Assessment Layer describes the types of assessment performed; and Layer IV: the Assessment Process Layer describes how the assessed data are processed further for the evaluated properties.

C. The Data Collection and Observer Layer

From a technical perspective, we classify whether these data in the Data Collection and Observer Layer (Layer II) as 1) are automatically retrieved and processed, e.g., log files, technical parameters, event lists, sensor data, or physiological data; 2) are data from surveys and questionnaires; these data are often coded and analysed after the visitors have left the site, and the answering process might be intrusive; 3) are observations by an external observer; or 4) are static data that are stored, available, or known, e.g., from databases, or historical data.

D. The Assessment Layer

For defining the categories used in the Assessment Layer (Layer III), we adapt the assessment categories presented by Leister and Tjøstheim [19] into the following components: *a) subjective assessment* based on questionnaires and ratings; *b) objective assessment* based on measurements at the object; *c) physiological assessment* based on sensor data from a subject; *d) behaviour and interaction assessment* based on observations of the subject and the subject's behaviour and interaction with other visitors; and *f*) objective and subjective context information, including visitor type.

E. The Assessment Process Layer

The Assessment Process Layer (Layer IV) describes how the data from the Assessment Layer are processed. In Figure 2, the impact of these data is shown with bold arrows. Additionally, values with dashed lines could be taken into consideration. Data that are visualised with dotted lines are used in the calibration process when creating the estimation model or for evaluation purposes. Most of these data cannot be automatically processed and need human intervention of some kind.

Layer IV contains the following elements:



Figure 2: Four-layer assessment framework for engagement of visitors using installations in science centres and museums.



Figure 3: The VEI profile for selected installations in three science centres.

1) Estimation Model: The estimation model is a mathematical model that takes measurable assessment data as input and returns estimated values expressed in suitable metrics. The estimation model usually returns an estimated value for one subject at a time since personal data specific to the subject are involved in the calculation. Machine learning approaches [18] can be used to implement the estimation model.

2) Collective Assessment: Collective assessment presents the rating for one installation based on the individual assessments by many subjects.

3) Measures for evaluated properties: The result of the assessment process consists of measures for the evaluated properties. This can be a vector of values that will be used in the process that requires such assessment data.

IV. THE VISITOR ENGAGEMENT INSTALLATION (VEI) PROFILE

To characterise installations, we developed the VEI profile in an iterative process with three sciences centres: the Engineerium (ENG), the Norwegian Museum of Science and Technology (NTM), and the Norwegian Maritime Museum (NMM).

Most studies that evaluate installations in science centres evaluate the impact of one dimension, such as interactivity, on the visitor. For this, observations of visitors are performed with various degrees of the dimension in question. However, we did not find a profile that characterises installations in multiple dimensions directly from an objective perspective, i.e., from only evaluating the installation.

The VEI profile was developed from a set of requirements for a well-working installation given by the participating science centres. From these requirements, we selected a set of dimensions that we considered sufficiently orthogonal and tried these on a set of fourteen selected installations (see Figure 3). We performed several iterations of this process until the requirements for common science centre installations were covered. We are aware that other dimensions could have been used. If necessary, our profile can be extended with more dimensions, such as immersion or degree of difficulty.

A. Defining the VEI Profile

The VEI profile classifies installations in their dimensions of competition (C), narrative (N), interaction (I), physical (P), visitor (user) control (U), and social (S). Each of these dimensions can have a value from 0 to 5; the higher the value, the more a dimension is present in an installation. TABLE I presents the description of the values for each dimension.

The dimensions of the VEI profile are described as follows:

[0	1	2	3	4	5]
С	visitor observes only; no competition element.	inst. has several components; result must be achieved to proceed or succeed.	visitor receives a score; competition with the installation (machine).	competition with other visitors asynchroneously.	competition with other visitors in real-time.	challenge in team; influence on other players' result.	С
N	no narrative; object can only be ob- served.	installation is used in a specific se- quence; chronolog- ical succession of events.	installation is built up in sequences; conditions must be met to proceed to next phase.	installation designed for multiple visitors; visitors may cooper- ate; multiple parallel narratives.	multi-player game or simulation; visitors cooperate to achieve a final result.	visitor develops nar- rative.	N
I	no interaction with object; observe only.	primarily no interac- tion; visitor can do something with the installation.	some interaction, such as "continue", "stop", "yes/no"; installation reacts.	moderate degree of interaction; choices influence outcome.	high degree of inter- action; choices have consequences; con- tent is stored.	visitor creates some of the content.	I
Р	no physical activity; observation only.	push buttons; touch screen; hold or touch object.	visitor moves betw. parts of installation; enter installation; guided tour.	some activity, e.g., operating pumps; throwing balls.	full body-motion; longer physical activity.	full body motion over time; performing physical task in real setting.	Р
U	controlled; visitor is observer; linear structure.	controlled with some degrees of freedom; mostly linear structure.	combination of controlled and free flow; choices can be made.	visitor can make choices; receives feedback on right or best choices.	visitor controls flow, but installation lim- its choices.	visitor has high de- gree of control; cre- ative process.	U
s	single visitor.	single visitor, others observe.	several installations used independently from each other.	single visitor while others observe and engage and cheer.	installation intended for several simulta- neous visitors.	multi-visitor instal- lation; visitors must cooperate.	s
	0	1	2	3	4	5	

TABLE I: EXPLANATION OF THE VALUES USED IN THE VEI PROFILE.

1) Competition: the degree of competition in an installation.

2) *Narrative:* the degree of active participation in the underlying narrative.

3) Interaction: the degree of interaction between the visitor and the installation.

4) *Physical:* the degree of physical activity the visitor must perform when using the installation.

5) *Visitor control:* the degree a visitor can control the use of the installation.

6) Social: the degree of social interaction between visitors.

B. Applying the VEI Profile To Measure Engagement

We applied the VEI profile to installations from the three science centres: five at ENG, five at NTM, and four at NMM. The VEI profiles of these installations are shown in Figure 3.

We assessed installations with visitors. We wanted to determine whether a change in one dimension of the VEI profile from x to y will result in a change of the visitor's engagement. For example, the assumption that a change in an installation with a C-factor (competition) of 3 to 4 would increase the visitor engagement could be tested by measuring the visitor engagement with the originally designed installation, make changes in the installation to increase the C-factor (e.g., making the competition with other visitors happen in realtime), and then measure the visitor engagement for the altered installation. We are interested in the relative changes of the assessed engagement-related values when testing installations with modified versions that have a different VEI profile.

C. Characterising Exhibitions Using the VEI Profile

Besides single installations, the VEI profile can be used to characterise exhibitions or groups of installations. For example, the graphical representation of the VEI profile for selected installations in Figure 3 suggests that physical activity is characterised as low for these installations. Also the Ndimension seems to be low, with the exception of two recently developed installations that are based on longer narratives. We also observe differences between the three sites regarding their overall profile characterised by mean values and variance of the respective VEI profiles.

V. ASSESSMENT OF A SELECTED INSTALLATION

We are doing assessments to analyse the correlations between the various data in Layer III of our assessment framework. These assessments will be used to build the structure and parameters of the estimation models in Layer IV.

In the following, we present preliminary results of an assessment that has the assumption that the C-dimension of the VEI-profile has an impact. We compare subjective data of winners, losers, and single players of a quiz game.

A. Experiment Setup

The installation *Footprint eQuiz* at the Engineerium, here denoted as ENG-12, shall challenge the visitors with questions about different environmental perspectives, show how the oil and gas industry takes responsibility, and how they work to minimise the negative impact on the environment. The installation provides an understanding of different ways we can lower our energy consumption to reduce the environmental impact.

ENG-12 is a game where up to two players compete by answering questions related to energy and the environment. There are two levels available, beginner and expert. The installation consists of two stations with two large buttons each, an orange one and a blue one. ENG-12 starts with a short introduction before ten questions are shown on the screen in sequence. As a question is shown, a timer starts counting down



Figure 4: VEI profiles of the ENG-12 installation when two players compete (solid line) and a single player version (dashed line); the single player version has lower values for C and S.



Figure 5: The installation ENG-12 at the Engineerium during the assessment.

to zero. Players answers by pressing either button within the countdown time. Players receive points for a correct answer and bonus points based on how quickly they answered. A player answering incorrectly loses points but can't go below zero. After the ten questions, a summary with the number of points scored for each player is presented.

In terms of the assessment model, the Scenario Layer contains the installation as the artefact (object) under observation, the visitors are the subjects, and the main action is to answer the questions by pressing buttons. The group of other visitors is the peer player. In the Data Collection and Observer Layer, we observe technical parameters from the installation, use a face reader and human observers to interpret emotions, and use surveys. Thus, in the Assessment Layer, technical parameters, physiological responses, and subjective assessment are employed. Since we are early in our investigations, the Assessment Process Layer is not yet fully implemented.

Figure 4 shows the VEI profile of ENG-12 with the solid line. We also show a version where only one player answers questions with the dotted line. This change lowers the values of both the C-dimension and the S-dimension.



Figure 6: Response scores on a Likert scale for winners (n = 29, mean game score: 1966), losers (n = 30, mean game score: 1354), and single players (n = 6, mean game score: 1837) for the subjective constructs Q16 (too difficult), Q20 (engagement), Q4 (intention to answer again), Q13 (fun), Q10 (concentration), and Q18 (intention to learn more).

Figure 5 shows the installation ENG-12 during the assessment. In addition to the installation, we have installed two cameras that observe each of the players, one camera that observes the scene from behind, and, for each player, a human observer makes notes. The video footage is used both for manual analysis and automated analysis of facial expressions using the Face Reader software by Noldus [20]. We also made changes to the installation's software to log all events (e.g., which button is pressed, and score values) with timestamps.

The observers note visitor's mood using a simplified valence tracker [21], i.e., whether the visitor is excited-positive, excited-negative, or calm-neutral for each quiz question. These values are compared with the outcome of the Face Reader software. The self-reported data by the visitors consist of a self-developed questionnaire for ENG-12 and a 20-item PANAS scale [22]. Since we are interested in the the positive affect, i.e., the PA of the PANAS, we omitted factors that express negative emotions (e.g., *guilty* or *scared*) that hardly can be an impact from the use of the installation.

We performed tests to ensure that the preliminary technical setup is in place and working. This includes logging the events from the installation (objective data), interpretation of the video footage and light conditions, usefulness of the questionnaires and valence tracker, and conformance with the Norwegian privacy laws. Still, challenges need to be addressed, such as lighting problems or adjustments in the questionnaires (some items of the PANAS adjectives seem not to be understood by the target group; as a consequence, we did not use these items).

B. Results

We asked students from school classes that visit the Engineerium to use ENG-12 with our assessment equipment and observed them as described above. In four sessions between October 2014 and January 2015 we assessed data from 29 winners, 30 losers, and 6 single players. The data from one of TABLE II: PANAS SCORES FROM THE EXPERIMENT.

PANAS	Pos.	Neg.
winners $(n = 29)$	34.0	16.5
losers $(n = 30)$	31.5	18.5
single players $(n = 6)$	34.0	20.3
std. dev. $(n = 65)$	6.8	5.0

the winners was discarded due to an irregularity (he played the game twice). We are aware that the number of single players is too low to give a significant result, and one of the single player responses is an outlier. So, we refrain from interpretations of the single player data. We show results from the subjective answers the players gave after having played ENG-12 with six selected questions in Figure 6. In TABLE II, we show the mean values of the positive and negative PANAS scores for the three groups and the mean value. We note that the standard deviation is in a similar range as published by Watson et al. [22] for assessments in the moment.

In our experiments, the automated face expression recognition fails in about 50% of the cases. The reasons for these failures include lighting problems (the light settings in science centres are often problematic for such analysis) and positioning of the cameras (these should be installed so that they do not obstruct essential parts of the installation). Given the achievable data quality of the data sets (14 winners and 17 losers), we registered about 70% smiles when an incorrect answer was given and about 40% smiles when a correct answer was given, independently whether they turned out to be winners or losers. Note that the smiles occur before the players know their ranking (winner or loser). The data from the valence tracker were only used to verify whether the assessment from the face reader is viable.

C. Interpretation

The interpretation of these data show rather small differences between winners and losers. However, a trend is visible: losers find the quiz questions somewhat more difficult (Q16). While they show lower engagement (Q20), their intention to answer again (Q4) and to learn more (Q18) is higher. They also report less fun (Q13) and less concentration (Q13). The PANAS scores show a similar trend, i.e., winners have a higher positive score while losers have a higher negative score. Note, however, that the differences are rather small. We also note that the trends in these responses are as expected between winners and losers. The data for the single players are not as expected, but due to low data quality we refrain from an interpretation.

For evaluating the impact the C-dimension in the VEI profile to the QoE, we do not yet have sufficient data quality, specifically for the single players. The fact that winners and losers show different values in the expected manner, both for the questionnaire and for the PANAS, shows that the C-dimension has an impact; else the two groups would not have shown differences.

The result concerning the number of smiles after each question suggests that the smiles might have a different social functionality than expressing enjoyment. However, the high number of smiles, specifically when answering incorrectly, show that the visitors are engaged and show emotions; that is that they are not indifferent. This also shows that it, in fact, is feasible to register engagement automatically.

VI. CONCLUSION

We presented the VEI profile to characterise installations at science centres, and a framework for assessing visitor engagement for installations. The goal is to assess engagement using measurable values from the installation, sensors, cameras, and so on, instead of using long-term observations and interpretation methods. Our current work shows the principles how to achieve this goal.

Currently, we have performed some preliminary assessments with ENG-12 with the metrics described here. The experiments so far have shown that registering engagement automatically is feasible. We need to perform more assessments with ENG-12 to get better data quality, as well as assessing other installations, the impact of other dimensions of the VEI profile, and the measurement of other data types in Layers II and III of our framework. While the goal is to develop a suitable estimation model in Layer IV of our framework, the collected data are not yet sufficient to apply machine learning methods.

VII. ACKNOWLEDGMENTS

The work presented here has been carried out in the project VISITORENGAGEMENT funded by the Research Council of Norway in the BIA programme, grant number 228737. The authors wish to thank Hege Røsto Jensen for her involvement in the assessment process.

REFERENCES

- P. Pierroux and A. Kluge, "Bridging the extended classroom: Social, technological and institutional challenges," Nordic Journal of Digital Literacy, vol. 6, no. 3, 2011, pp. 115–120.
- [2] A. Hofstein and S. Rosenfeld, "Bridging the gap between formal and informal science learning," Studies in Science Education, vol. 28, no. 1, 1996, pp. 87–112.
- [3] J. H. Falk and L. D. Dierking, "Learning from the outside in," in Lessons without limit: How free-choice learning is transforming education. New York: Altamira press, 2002, pp. 47–62.
- [4] J. Wang and A. M. Agogino, "Cross-community design and implementation of engineering tinkering activities at a science center," in Proc. FabLearn 2013: Digital Fabrication in Education, Stanford, 27-28 October 2013, pp. 1–4. [Online]. Available: http://fablearn.stanford.edu/2013/papers/ [Accessed: 2 Feb 2015].
- [5] M. Lindauer, "What to ask and how to answer: a comparative analysis of methodologies and philosophies of summative exhibit evaluation," museum and society, vol. 3, no. 3, November 2005, pp. 137–152.
- [6] A. Šuldová and P. Cimler, "How to assess expirience the new trend in research technique, use in nonprofit sector of entertainment and educational industries," Marketing a obchod, vol. 4, 2011, pp. 115–124.
- [7] C. W. Sanford, "Evaluating Family Interactions to Inform Exhibit Design: Comparing Three Different Learning Behaviors in a Museum Setting," Visitor Studies, vol. 13, 2010, pp. 67–89.
- [8] L. D. Dierking and J. H. Falk, "Redefining the museum experience: the interactive experience model," Visitor Studies, vol. 4, no. 1, 1992, pp. 173–176.

- [9] J. Falk and M. Storksdieck, "Using the contextual model of learning to understand visitor learning from a science center exhibition," Science Education, vol. 89, 2005, pp. 744–778.
- [10] J. H. Falk, "The impact of visit motivation on learning: Using identity as a construct to understand the visitor experience," Curator, vol. 49, no. 2, 2006, pp. 151–166.
- [11] J. H. Falk, J. Heimlich, and K. Bronnenkant, "Using identityrelated visit motivations as a tool for understanding adult zoo and aquarium visitors' meaning-making," Curator: The Museum Journal, vol. 51, no. 1, 2008, pp. 55–79.
- [12] C. Barriault and D. Pearson, "Assessing Exhibits for Learning in Science Centers: A Practical Tool," Visitor Studies, vol. 13, 2010, pp. 90–106.
- [13] D. L. Young, "A phenomenological investigation of science center exhibition developers' expertise development," Ph.D. dissertation, The University of North Carolina at Chapel Hill, 2012, 174 pages.
- [14] S. Allen, "Designs for learning: Studying science museum exhibits that do more than entertain," Sci. Ed., vol. 88, no. S1, 2004, pp. S17–S33.
- [15] D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card, "The cost structure of sensemaking," in Proc. of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. New York: ACM, 1993, pp. 269–276.

- [16] G. Klein, B. Moon, and R. R. Hoffman, "Making sense of sensemaking 1: Alternative perspectives," IEEE Intelligent Systems, vol. 21, no. 4, Jul. 2006, pp. 70–73.
- [17] —, "Making sense of sensemaking 2: A macrocognitive model," IEEE Intelligent Systems, vol. 21, no. 5, Sep. 2006, pp. 88–92.
- [18] C. M. Bishop, "Pattern Recognition and Machine Learning (Information Science and Statistics)". Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [19] W. Leister and I. Tjøstheim, "Concepts for User Experience Research," Norsk Regnesentral, NR Note DART/14/2012, 2012.
- [20] V. Terzis, C. N. Moridis, and A. A. Economides, "Measuring instant emotions during a self-assessment test: The use of facereader," in Proc. Measuring Behavior 2010, A. J. Spink, F. Grieco et al., Eds., Eindhoven, The Netherlands, 2010, pp. 192–195.
- [21] G. Chanel, K. Ansari-Asl, and T. Pun, "Valence-arousal evaluation using physiological signals in an emotion recall paradigm," in Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on, oct 2007, pp. 2662–2667.
- [22] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the PANAS scales," Journal of Personality and Social Psychology, vol. 54, 1988, pp. 1063–1070.