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, Andreas Larssen, Michel de Brisis
Expology
Oslo, Norway
{goran, andreas, michel}@expology.no

***Abstract*—Science centres and museums struggle to measure how engaging specific installations are for visitors. We present a framework for assessing visitor engagement by using non-intrusive technologies. We present a profile for mapping out an installation over eight dimensions and how we created it. We also present techniques for performing assessments using the facial expressions of visitors and asking short, targeted questions to visitors. Combining these together results in a fast assessment that happens as a visitor interacts with an installation. We have performed evaluations of three different installations in three science centres: one looked at the role of competition in exhibits, one at ways to automate the assessment, and another at how altering components like narrative in an installation affects the assessment's result. The assessment framework and classification method work in multiple installations and form the basis for a new tool for measuring engagement in a visitor centres and museums.**

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

I. Introduction

Science centres and museums present exhibitions, installations, and educational programmes that should engage visitors for self-education on a subject and to inspire the visitors to learn more. There is little data showing how well these installations perform in transferring knowledge to the visitors. Similarly, there is little data to determine whether modifying an installation increases a visitor's engagement. Previously, we proposed a concept for a system that can give evidence to these questions in real-time [1]. This concept was supported by a further study [2], and the current paper builds on this work.

Our objective is to measure the performance of installations, but we assume we cannot measure this directly. Instead, we assess the engagement of visitors while they use the installation and retrieve parameters and objective data from the installation and its context. We intend to avoid time-consuming observations by the museum staff and keep intrusive methodologies, such as questionnaires, to a minimum.

We argue that we can assess dimensions of engagement in an installation using 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 adjusting these installations contribute to a better engagement and experience.

Observations by museum personnel tend to focus on the visitor instead of the installation. Common methods for collecting data from visitors include interviews and questionnaires. But long interviews or questionnaires might be intrusive for the visitor, and the answers are given in retrospect, i.e., not *in situ*. Our approach is to use observations from sensors to retrieve data about a visitor's engagement. Electronic questionnaires will be tailored so that only relevant questions will appear. Thus, the visitor will not be bothered more than absolutely necessary.

First, we present an overview of related work, showing 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 eight dimensions (Section IV). An assessment of selected installations follows (Section V). Finally, we present our conclusion (Section VI).

II. Related Work

It is well-documented that science centres and museums perform visitor studies. While demographic data about visitors and information on their enjoyment is assessed, the impact of these data is more difficult to grasp [3, p.169]. Many science centres have performed visitor studies to varying degrees. These studies include statistical data from ticket sales, questionnaires, feedback from visitors, observations, and larger studies [4] [5] [6]. Several studies focus on the learning aspect. Other studies evaluated whether the visitors enjoyed their visit, whether an installation works as intended, and how installations could be changed for a better presentation of the contents.

Science centres are informal learning environments [7] that are distinct from classrooms because they offer free-choice learning [8][9], i.e., visitors can choose which activities to participate in and they can leave at any time. Visitor studies have been performed since the late nineteenth century. In 1884, Higgins [10] mentions that observations of visitors and asking them for remarks might lead to valuable information.

Lindauer [11] presents a historical perspective of methodologies and philosophies of exhibit evaluations. Lindauer lists 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 deal with the assessment of learning, often using a longitudinal approach [12], i.e., observing a subject or installation over time. Šuldová and Cimler [13] suggest that engagement can be assessed more instantaneously and be used as a part of learning assessment, supporting Sanford's [14] claim that "some compelling evidence links visitor engagement to learning".

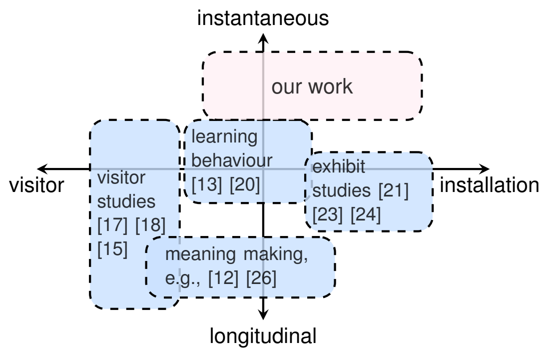


Figure 1: Classification of selected work in visitor studies

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

According to McManus [15], the visitor instead of the artefact has been the focus in visitor studies since the 1980s. These visitor studies include demographic characteristics and segmentation, behavioural and knowledge gain studies, and visitor focused studies. Yalowitz and Bronnenkant [27] give a review of methodologies for timing and tracking visitors in exhibitions, also giving advice on how to perform assessments of visitor behaviour. Various methodologies have been developed to examine the behaviour of visitors in museums in detail [28] [29] [30] [31].

Dierking and Falk [16] present the Interactive Experience Model, which is a visitor-centric model. They define the interactive experience influenced by three contexts: *1*) the personal context, *2*) the physical context, and *3*) the social context. Falk and Storksdieck [17] use the principle of identity-related motivation to place visitors into five identity types: *1*) the explorer; *2*) the facilitator; *3*) the professional and hobbyist; *4*) the experience seeker; and *5*) the spiritual pilgrim. Variables, such as prior knowledge, experience, interest, visitor agenda, and social group are encapsulated in these identity types. This line of research has been further studied [18][19].

Barriault and Pearson [20] present frameworks that analyse the learning experience near instantaneously by identifying learning-specific behaviour observed by cameras and microphones installed within an installation. Šuldová and Cimler [13] refine these methods, but still depend on manual analysis.

Recently, Pierroux and Steier [32] presented the Visitracker, a tablet-based system for registering the visitors' behaviours. The graphical interface replaces the manual note-making, but a human observer is still needed to register the events.

In longitudinal visitor studies, observations and sense-making [26] are often used. In sense-making, qualitative mental models, understanding events, and an iterative approach for interpretation of situations (e.g., the data/frame theory of sense-making [33][34]) are in the foreground. But 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.

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.

Shettel et al. [21] present a more installation-centric approach where they evaluate exhibits by means of visitor observations and questionnaires using the technology available at the time, such as video tape recordings. They observe how visitors behave toward installations to determine how effective an exhibit is.

Alt and Shaw [22] present a study where visitors characterise installations using a list of phrases, both positively and negatively loaded. The phrases mentioned most often are then compared with the goals of the museum to identify where the installations can be improved.

Spegel [23] presents the *Expogon*, a graphical classification used as a mind map for exhibit planners when going through a museum. Note that the purpose of the Expogon is to stimulate and inspire on a subjective (qualitative) basis rather than to measure. The Expogon breaks down the exhibition medium into six elements: *1*) narrative, *2*) space, *3*) visitor, *4*) objects, *5*) time, and *6*) sender. Each element consists of fifteen hexagons representing categories, ten pre-filled and five empty for additional categories. The researcher wanders through an exhibition and notes observations on the Expogon. Thus, it is a qualitative tool that allows brainstorming when evaluating an exhibition. The Expogon gives hints to an evaluator on what to improve in an exhibition. However, it does not reflect to what degree the six elements are fulfilled. To rectify that, we developed a different approach with the VEI profile [1] that we extended in the current paper.

Young [24] suggests that developers need to advocate for the visitors and think as a visitor; Young recommends a cyclical development process. Allen [25] presents a study of three different versions of an exhibit for the purpose of studying dimensions of interactivity.

C. Observation Methodology

Traditionally, visitor studies use assessments where observers are placed near the installations. These observers make notes of the visitors' actions related to their use of the installations. Methods include counting and making notes. The visitors are often asked to fill out questionnaires related to their visit. However, such methods are often perceived as being intrusive and, thus, can reduce the visitor's experience.

Tröndle et al. [35] show an innovative possibility of combining movement tracking, physiological data (heart rate, skin conductivity, etc.), and psychological data. Any single senor for measuring has weaknesses and limitations when used in visitor studies [14]. However, when using multiple sensors concurrently the result will become better, provided that the method and the weight factors in an estimation model are calibrated correctly. Oppermann [36, p.145] posits that “a multi-method approach allows researchers to be more confident about their results”.

We found few examples where physiological data for studies in science centres were used [35, 37]. Other data sources are mimics from video- and image data, prosody in sound data (e.g., intonation, pitch, strength), gesture recognition, or other bio-physiological sensor data.

For semi-automatic data assessment of visitor engagement, some components, such as face recognition and emotion recognition, are already available [38, 39]. Others have studied how to assess visitors’ physical reaction using galvanic skin response in art exhibitions [37]. It is known that observations, e.g., visual or auditive recordings, and physiologic data measured with sensors are correlated with visitor experience; however, there is no unambiguous correlation.

Our assessment methodology is inspired by the model described by Russell [40], which shows a relation between psycho-physiological reactions and emotions. O’Brien [41] posits that engagement has been defined as *to involve the user emotionally when interacting with a system.* Engagement can be quantified by *focused attention* [42], measured, e.g., by using questionnaires after a museum visit or after the use of an installation. A modified version of the *Menorah Park Engagement Scale* together with the observation tool by Šuldová and Cimler [13] can be used for observing and classifying into categories [42]. See also the work by Barriault and Pearson [20], Griffin [43], and Griffin and Paroissien [44].

Wilhelm and Grossmann [45] and Nacke and Lindley [46] have shown the connection between emotions and psycho-physiological reactions, such as skin conductance, breath (strength, frequency), ECG, and EEG. These reactions has been used systematically in *quality assessment* studies [47].

Picard [48] coined the term *affective computing* to describe using computers and sensors to interpret emotions. Hoque et al. [49] show examples where facial expressions in camera images are used to interpret emotions. Ben Ammar et al. [50] have shown adaptive systems, e.g., within learning. Emotion recognition is a field where expressions can be interpreted (e.g., facial expressions, gestures, movements, voice) or physiological reactions (e.g., skin conductance or changes in the face colour [51]). A recent research challenge is the interpretation of multi-modal expressions.

Witchel et al. [52, 53] posit that non-instrumental movement inhibition can be used as a manifestation and proxy for engagement. According to them, cognitive engagement is an embodied phenomenon that can attenuate certain types of non-instrumental movements, such as larger postural movements and self-adaptors. They also found that non-instrumental movement disinhibition can be an indicator of engagement, e.g., during breaks, or disengagement, e.g., when occurring during active parts of a presentation. Their experiments were performed for screen-based visual stimuli. Research is needed to evaluate whether their arguments apply to visitors in science centres and museums interacting with installations.

D. Engagement and Gamification

Gamification is the application of game-design elements and game principles in non-game contexts [54, 55] to improve user engagement, productivity, learning, flow, etc. Kapp [56] presents gamification in the context of learning. We argue there are similar opportunities for using installations and gamification elements when it comes to learning.

Dixon [57] gives a brief history of the concept of player types, starting with Bartle’s [58] concept of four player types. Marczewski [59] presents user types for game players, comparable to the classification by Falk and Storksdieck [17] for installations. Marczewski classifies users into *philanthropists*, *achievers*, *free spirits*, *socialisers*, *disruptors*, and *players*; he explains their roles and preferences.

Legault [60] presents twelve gamification elements that apply to e-learning. These elements are: *1*) narrative (story, protagonist, antagonist, plot), *2*) rules, *3*) player control, *4*) discovery and exploration possibilities, *5*) interactivity, *6*) feedback (provide a cue to player about progress), *7*) time constraints (create a sense of urgency), *8*) loss aversion (humans prefer avoiding losses; loss is twice as powerful as a gain), *9*) continuous play (after interruption), *10*) reward, *11*) levels (achieving different levels, goals, or challenges), and *12*) competition.

Hamari et al. [61] presents a literature review of empirical studies on gamification. In this context, they classify literature that refers to motivational affordances into ten categories: *1*) points, *2*) leaderboards, *3*) achievements and badges, *4*) levels, *5*) story and theme, *6*) clear goals, *7*) feedback, *8*) rewards, *9*) progress, and *10*) challenge. Weiser et al. [62] present a taxonomy of motivational affordances for meaningful gamified and persuasive technologies. Their taxonomy comprises of elements (including assignments, achievements, leaderboards, reminders, points, virtual goods, friends), mechanics (including feedback, rewards, education, competition, challenges, cooperation), and general design principles (including personalise experiences, offer meaningful suggestions, support user choice, respect stages of behaviour change, and provide user guidance). In our current paper, we will consider suitable gamification elements to extend our previous work.

III. Approach

Museums and science centres are places where visitors learn and gain knowledge through encountering and engaging with installations. These installations 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.

Our methodology is illustrated in Figure 2. We start by selecting the characteristics of installations. Currently, we use the VEI profile (Section IV) to identify potential characteristics

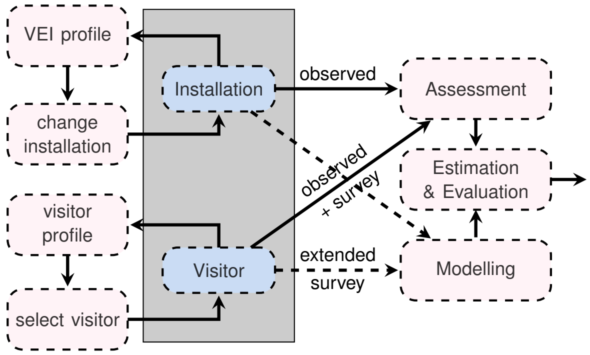


Figure 2: Method applied in our research to evaluate properties of installations

along eight dimensions that could be altered to achieve a design goal for the installation. The success of these changes can be evaluated, and the results can be compared.

Whether a goal is achieved can be assessed using the assessment framework presented in Section III-A. Additionally, the visitor characteristics, e.g., the characteristics described by Falk and Storksdieck [17], are used for the selection of visitors that act as respondents in the assessment.

The evaluation process for an installation, shown on the right side of Figure 2, consists of a *modelling phase* and an *assessment phase*. In the modelling phase, we establish an estimation model that allows us to estimate values and parameters from only few available data. If we succeed with this, it will be sufficient in the assessment phase to retrieve few data and use only a few questions to the visitor to extract a rich body of information about the installation.

We use all available data about the installation and the visitor in the modelling phase to establish the estimation model (stippled lines in Figure 2). In current work, we use statistics and regression analysis to establish correlations between data, but we intend to use a machine learning approach [63] to establish more complex models later.

In the assessment phase, we use the previously established estimation model. We gather information about visitors' interaction with the installation using diverse data, observation, and responses from selected questions to estimate a goal (e.g., how satisfied or engaged a visitor is during a visit).

When performing the assessment we need to find the minimum number of sources to provide a valid score. The goal is to make the assessment as non-intrusive as possible. Questionnaires may be used, but the questions that are asked are targeted and will only be about the installation or an aspect of the installation. Using this methodology, museums can then change the installation, run a new assessment and see the effects of the change in the installation's VEI profile.

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. A generalised version of this assessment framework is presented by Leister [64].

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.

Our assessment framework (Figure 3) 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.

B. 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*) data automatically retrieved and processed, e.g., log files, technical parameters, event lists, sensor data, or physiological data; *2*) 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*) data from observations by an external observer; or *4*) static data stored, available, or known, e.g., from databases, or historical data.

C. The Assessment Layer

For defining the categories used in the Assessment Layer (Layer III), we adapt the assessment categories presented by Leister and Tjøstheim [65] into the following components: *a*) subjective assessment based on questionnaires and ratings, *b*) objective assessment based on measurements of the artefact, *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 both the object and other subjects, *e*) observation of the subject and interaction with other visitors, and *f*) objective and subjective context information – including visitor type.

D. The Assessment Process Layer

The Assessment Process Layer (Layer IV) describes how the data from the Assessment Layer are processed. In Figure 3, 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:

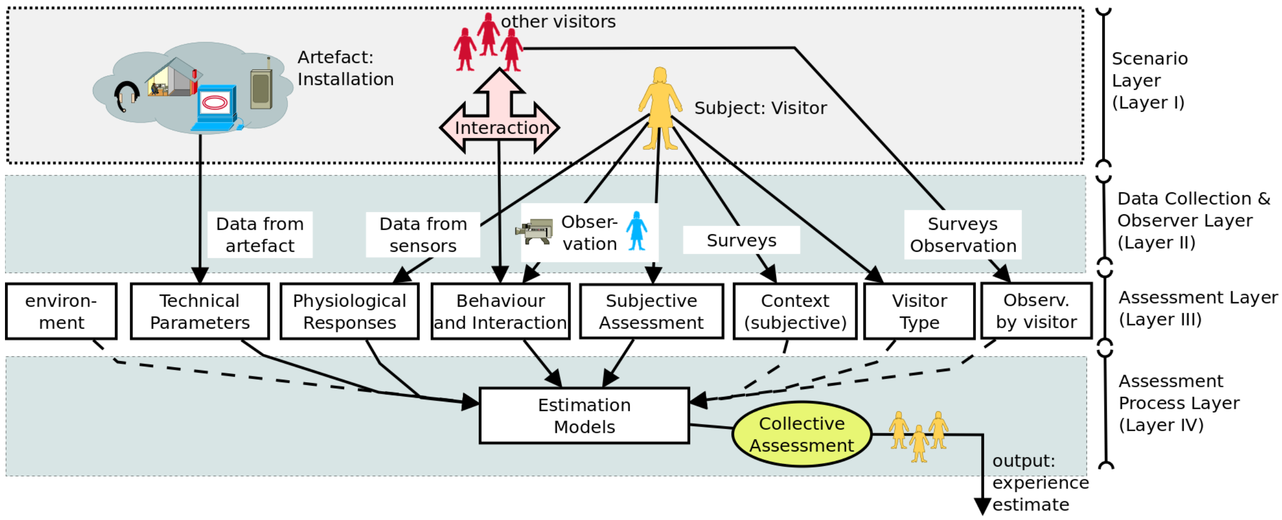


Figure 3: Four-layer assessment framework for engagement of visitors using installations in science centres and museums.

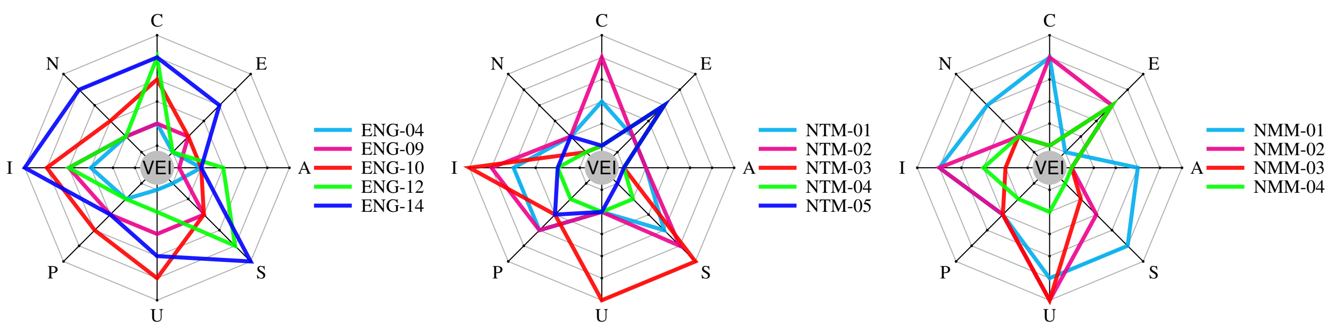


Figure 4: The VEI profile for selected installations in three science centres.

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 [63] 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. Characterising Installations

Installations can have a variety of qualities and characteristics. The design of these installations is important for the engagement and experience outcome for visitors. However, the assessment of the installation design is often unstructured. To develop a more structured way of quantifying the characteristics in an installation, we developed the Visitor Engagement Installation (VEI) profile that can assess properties of installations and give hints for the designers on how to improve the experience.

To characterise installations, we developed the original VEI profile [1] 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).

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 4). These installations range from simple vitrine exhibits to complex games or simulations where several visitors compete.

We performed several iterations of the finding process for the set of dimensions until the requirements for common science centre installations were covered.

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.

A. The current Visitor Engagement Installation (VEI) Profile

The VEI profile characterises installations. The original defined by Leister et al. [1] had six dimensions, but we saw a need to extend the profile to make it more relevant for installations. We made adjustments to some of the definitions, e.g., the narrative has been adapted from its previous definition to findings from the literature [66]. We retrieved candidate dimensions from the gamification literature [56] [60] [61] [62]. This list is shown in TABLE I. We excluded dimensions that are related to already existing dimensions. This results in our current version with eight dimensions:

1) *Competition* (C): the degree of competition in an installation. It ranges from no competitive elements, to competition as a single player with high scores, to competing concurrently with other players.
2) *Narrative* (N): the degree the installation's narrative impacts the visitor. It ranges from no narrative (simply observing an object) to a fully developed, dramatic narrative with a story arc and characters.
3) *Interaction* (I): the degree of interaction between the visitor and the installation. It ranges from no interaction, to making choices that have consequences, to visitors creating their own content.
4) *Physical* (P): the degree of physical activity the visitor must perform when using the installation. It ranges from observing, i.e., no significant physical activity, to full body motion over time in realistic settings.
5) *Visitor (user) control* (U): the degree a visitor can control the use of the installation. It is also characterised by the size of the possibility space of user interactions. It ranges from no control over the installation (e.g., you can only read things in one order) to the ability to control the flow of the installation and add to its possibility space.
6) *Social* (S): the degree of social interaction between visitors. It ranges from a design for one visitor only, to groups of visitors interacting by themselves, to visitors that need to cooperate together to use the installation.
7) *Achievements* (A): the degree a visitor needs to be aware of achievements when using an installation. It is also characterised by the degree of feedback from the system. It ranges from no achievements to having a visitors achievements made concrete and the choices and their consequences displayed.
8) *Explore* (E): the degree of exploration or discovery for visitors in the installation. Exploration often can be done by trying out things with the possibility of failing, and thus learning from the failures. It ranges from predefined experiences to to exploring timeliness and possibilities spaces without penalty.

The new dimensions are *achievements* and *explore*. We also considered the dimensions such as *time constraints*, *cooperation*, *reminders*, and *challenges*, but did not find these most relevant for installations. The VEI profile is flexible and dimensions could be exchanged or more could be added depending on the purpose of a study.

TABLE I: Gamification elements and motivational affordances used to extend the VEI profile

#	Element	VEI	References	Comment
1	competition	C	[1] [60] [62]	
2	narrative	N	[1] [60] [61]	adjusted
3	interactivity	I	[1] [60]	
4	physical	P	[1]	
5	user control	U	[1] [60]	adjusted
6	social	S	[1]	
7	achievements	A	[67] [61] [62]	new
8	explore, discover	E	[60]	new
9	time constraints	T	[60]	considered
10	assignments, goals	(N)	[62] [61]	
11	challenges	(N)	[61] [62]	
12	feedback	(A)	[60] [62]	
13	rules	(U)	[60]	
14	loss aversion	(C,S)	[60]	
15	continuous play	–	[60]	
16	rewards	(A)	[60] [62]	
17	levels	(A)	[60] [61] [62]	
18	chance	(U)	[56]	
19	points	(A,C)	[61] [62]	
20	leaderboards	(A,C)	[61] [62]	
21	feedback	(A)	[60] [61]	
22	rewards	(A)	[61]	
23	progress	(A)	[61]	
24	reminders	(N)	[62]	
25	virtual goods	(A)	[62]	
26	friends, teams	S	[62]	
27	cooperation	(S)	[62]	

External influences are not taken into account in the VEI profile since these are not properties of the installation. Thus, physical factors, such as noise, light or smell need to be handled separately. We also exclude properties that belong to the context, such as social factors, institutional factors, or recent incidents personally or globally.

Each dimension has a value from 0 to 5; the higher the value, the more that dimension is present in an installation. TABLE II presents the description of the values for each dimension.

We posit that increasing each of these dimensions will potentially increase the visitor's engagement up to a point. At some point, the installation becomes too demanding or complex and the engagement will likely drop. There are implicit dependencies between dimensions for any installation, i.e., a change in one dimension may affect others. Where these points and dependencies are will depend on the installation.

B. Using the VEI profile to measure changes in 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 4.

We assessed installations with visitors. We wanted to determine whether a change in one dimension of the VEI profile would 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 real-

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

	0	1	2	3	4	5	
C	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.	**C**
N	no narrative elements added; object can only be observed.	non-dramatised story; explaining text only.	narrative structure with limited use of narrative or scenographic elements.	narrative structure with rich use of narrative elements in a scenographic setting.	a developed dramatised story with a narrative universe in a scenographic setting.	a visual, immersive environment with a strong dramatised narrative story.	**N**
I	no interaction with object; observe only.	primarily no interaction; 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 interaction; choices have consequences; content is stored.	visitor creates some of the content or develops narrative.	**I**
P	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.	**P**
U	controlled; visitor is observer; linear structure.	controlled; linear structure or chronological succession of events.	installation is built up in sequences; conditions must be met to proceed to next phase.	visitor can make choices, but the choices have no effect on the flow of the installation.	visitor controls flow, but installation limits choices; multiple parallel narratives.	visitor has high degree of control; creative 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 simultaneous visitors.	multi-visitor installation; visitors must cooperate.	**S**
A	no specific achievements possible with installation.	immediate feedback on failure or success; countdown timer.	achievement collected, only shown at the end.	current status represents achievements; progress bar; graphical visualisation.	achievements are shown; points, lists, gadgets are displayed.	achievements are shown; choices and their consequences are displayed.	**A**
E	installation allows pre-defined views only.	view from several perspectives.	explore while progress is stopped.	explore while progress is ongoing.	can dissect installation with recover-functionality.	can follow timelines or branches in possibility space.	**E**
	0	1	2	3	4	5	

time), 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. For some installations, a change in one dimension of the VEI profile might be too small to result in a significant change in engagement.

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 4 suggests that physical activity is characterised as low for these installations, as is the A-dimension. Also, the N-dimension 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.

Assuming that installations in an evaluation form one ensemble, we can visualise this ensemble's characteristics as a whole. In the above example, we recognise that the physical (P) dimension has rather low values for these installations. An exhibition designer could consider to increase the P-value by making changes to the installation for the sake of giving visitors a better experience or to decrease values, e.g., the U-dimension in the above example. In other cases, a good mix of characteristics could be the objective of an exhibition.

When an installation designer decides to make changes on the basis of the VEI profile, both the original and the modified installation need to be assessed with regard to how engaging these installations are. In the field of visitor studies several approaches are possible, such as observations, questionnaires, or assessment using diverse sensors. In the ideal case, the assessment is minimally intrusive, does not bother the visitor, and can be performed in a short time.

V. ASSESSMENT OF SELECTED INSTALLATIONS

We performed assessments to analyse the correlations between the various data and layers in our assessment framework. We did not aim at creating the complete estimation function, but to find evidence that it is feasible to create a function using properties and correlations.

We performed three assessments. In the first assessment, we assume that competition, i.e., the C-dimension of the VEI profile, has an impact on the visitor's engagement. Using a quiz game, we compared subjective data of winners, losers, and single players of a quiz game. When analysing the data, we received unexpected results with the interpretation of observed smiles: in a competition situation, visitors that answer wrong to an quiz question tended to smile more often than when answering correctly.

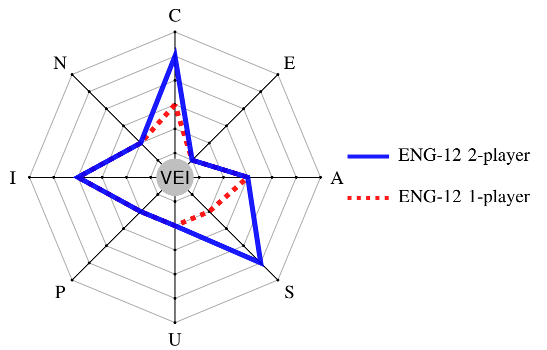


Figure 5: VEI profiles of the ENG-12 installation when two players compete (solid line) and a single player version (dashed line); the single player version has lower values for C and S.

Figure 6: The installation ENG-12 at the Engineerium during the assessment.

In the second, still ongoing assessment, we evaluated the impact of the C-dimension in a more complex game that lasts fifteen to twenty minutes. From the experiences of the first assessment, we tried to automate the assessment using events from the installation and using a Kinect to retrieve the players' emotions.

The third assessment looks at the influence of the narrative. For this, we used a type of installations where visitors toss balls on a wall without many explanations. After changing the narrative, we make similar evaluations with a different group of visitors and compare the results.

A. The Influence of Competition to Experience

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 to zero. Players answer by pressing either button before the timer reaches zero. Players receive points for a correct answer and bonus points based on how quickly they answered. Players lose points when answering incorrectly but the score cannot go below zero. After the ten questions, a summary with the number of points scored for each player is presented.

1) Experiment Setup: Figure 5 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 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 [68][69]. We also made changes to the installation's software to record all events (e.g., which button is pressed, points awarded, and player scores).

The observers note visitor's mood using a simplified valence tracker [70], 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 [71]. 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 needed 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).

2) 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 five sessions between October 2014 and March 2015 we assessed data from 33 winners, 34 losers, and 20 single players. All participants were between the ages of 14 to 16. The data from

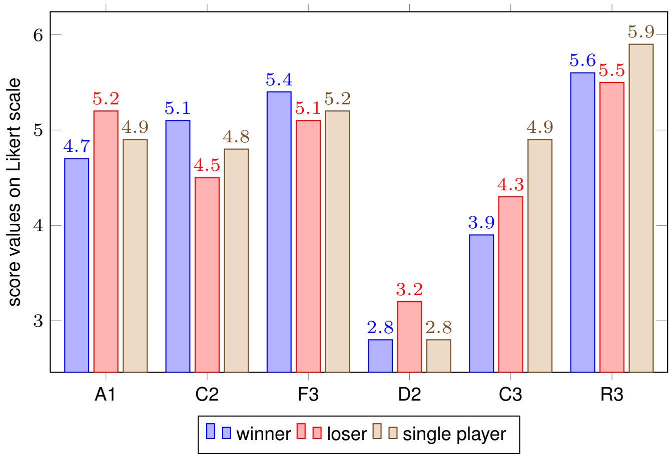


Figure 7: Response scores on a Likert scale for winners ($n = 33$, mean game score: 2009), losers ($n = 34$, mean game score: 1391), and single players ($n = 20$, mean game score: 1790) for the subjective constructs A_1, C_2, F_3, D_2, C_3, and R_3.

TABLE III: PANAS SCORES FROM THE EXPERIMENT.

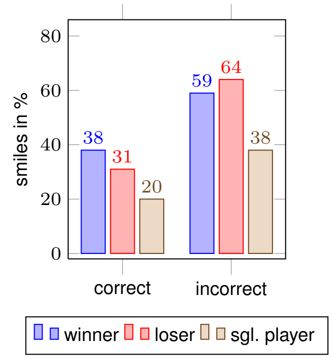
PANAS	Pos.	Neg.
winners ($n = 33$)	34.0	16.5
losers ($n = 34$)	31.5	18.5
single players ($n = 20$)	33.0	18.6
all ($n = 87$)	32.3	17.4
std. dev. ($n = 87$)	6.1	6.1

one of 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 7. In TABLE III, 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. [71] for assessments in the moment.

Figure 8 shows that we registered significantly more smiles when an incorrect answer was given than when a correct answer was given, independent of whether they ended up the winner or loser of the game after 10 questions. These smiles occur *before* the players know they win or lose the competition. We also observe that the number of smiles is significantly reduced for the single player games. We interpret this as a smile does not necessarily expresses happiness about answering correctly, as we first assumed. Instead, we need to re-interpret the smile to have a different function, e.g., a social function; this fact is supported from the field of psychology [72][73]. Yet the high number of smiles, specifically when answering incorrectly, show that the visitors are engaged and show emotions; they are not indifferent. This also shows that it is feasible to register engagement automatically.

The questions from the questionnaires is given in TABLE IV. The interpretation of the assessed data from these



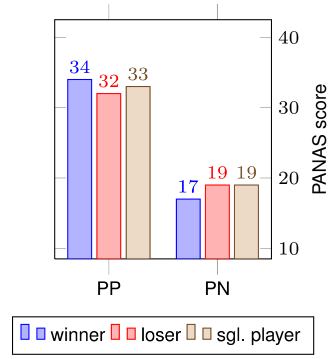


Figure 8: Smiles for correct answers and incorrect answers (left) and PANAS Positive and Negative values (right) for the three player categories winner, loser, and single player. While the differences for PANAS are small, there are significant differences for the number of smiles: visitors smile more often when answering incorrect.

TABLE IV: FORMULATION OF THE QUESTIONS FOR THE VARIABLES F_i, C_i, A_i, R_i, L_i, D_i, AND E, TAILORED FOR THE INSTALLATION ENG-12. ALL QUESTIONS ARE ON A SCALE FROM 1 (LOW) TO 7 (HIGH).

F_1	Did you have fun using the installation and answering the questions?
F_2	How much did you like the installation?
F_3	The installation was entertaining.
C_1	I concentrated so that I could answer as fast as possible.
C_2	I read the question. If I did not remember the text, I read it once more.
C_3	I was focused and carefully read the question. Then I made my judgement before I answered.
A_1	Do you want to use the installation one more time?
A_2	Do you want to use the installation a second time to improve your score?
A_3	At your next visit to the science centre, do you think you will use this installation?
R_1	Would you recommend the installation to others who are with you today?
R_2	I would like to recommend the installation to someone I know.
R_3	I will recommend the installation to other visitors to the science centre.
L_1	After reading a question, but unsure about the answer, was topic something that I would like learn more about.
L_2	After you had answered, did your interest in the subject increase or decrease?
L_3	The installation triggered my interest to learn more about energy and the environment.
D_1	How easy or difficult was it to answer the questions?
D_2	I thought the questions were too difficult.
E	How engaging was the installation?

questionnaires show small differences between winners and losers. However, a trend is visible: losers find the quiz questions somewhat more difficult (D_2). While they show lower engagement (R_3), their intention to answer again (A_1) and to learn more (C_3) is higher. They also report less fun (F_3) and less concentration (C_2). 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.

We cannot say for certain that VEI profile's C-dimension has an impact on the QoE because we do not have sufficient data yet. We need more data for single player games. Yet the fact that winners and losers have different values in the questionnaire and PANAS in the expected manner (that is winners have higher values than losers), shows that the C-dimension has an impact on two-player games. If it did not, the data from these two groups would be the same.

3) Automating the Value Tracker: In our experiments, we used both a valence tracker operated by observers and the FaceReader software by Noldus [68][69]. Both assessment methods have advantages and disadvantages. The valence tracker is a manual method performed by an observer where the opinion of the observer plays a role. But in the experiment settings, it is quite easy to make mistakes when registering emotions, e.g., having to focus on both players can make it difficult to capture the emotion from both before the start of a new question since the players normally are indifferent as they read a question. Thus, two observers were used in the experiments for ENG-12. The video footage can be used in the case of doubts, but this is time consuming. A human observer is visible for the visitors. Thus, for the experiment unwanted communication between visitor and observer can occur that might influence the result, also referred to as the Hawthorne effect [74]. In our studies, we have not taken this effect into account.

In our experiments, the automated face expression recognition fails in about half 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). Other problems occur when visitors temporarily turn their heads away or make hand movements that partially obstruct their face.

We discussed some of these issues with the developers of the FaceReader software. We found out that we ideally should have the camera at the same height as the head of the visitor rather than filming from a lower position. However, having the observation cameras at the ideal height might obstruct important parts of the installation. When placing cameras in future installations, the camera placement needs to be planned carefully; we also intend to evaluate whether competing technologies suffer from similar impacts.

4) Optimising Questionnaires: When assessing engagement and other properties, questionnaires are still necessary. However, to reduce intrusiveness of such questionnaires, we integrate the questionnaire into the natural flow of the visit and reduce the number of questions. To find out which questions are representative, we combined the positive-negative affect-instrument (PANAS) with a survey-instrument that captures the subjective experience of the player [2].

For this study, the evaluation used a within subject design, i.e., all students used eQuiz as a competition between two players. For the analysis, we used partial least squares (PLS), which is a structural equation modelling technique [75] that can simultaneously estimate measurement components and structural components (i.e., the relationships among these constructs). The PLS algorithm is a sequence of regressions

TABLE V: Summary of measurement scales for Research Model 1.

Construct	Measurement[a]	Factor Loading
Concentration composite reliability: 0.74	C_1	0.81
	C_2	0.42
	C_3	0.83
Enjoyment and engaging composite reliability: 0.88	F_1	0.68
	F_2	0.86
	F_3	0.87
	E	0.81
Intention to use again composite reliability: 0.88	A_1	0.85
	A_2	0.77
	A_3	0.89
Positive Affect composite reliability: 0.84	Active	0.72
	Excited	0.64
	Enthusiastic	0.77
	Inspired	0.67
	Attentive	0.78
Negative Affect composite reliability: 0.83	Nervous	0.77
	Afraid	0.77
	Frustrated	0.60
	Scared	0.69
	Upset	0.67

[a]The formulation of the measures for the questionnaires are presented in TABLE IV

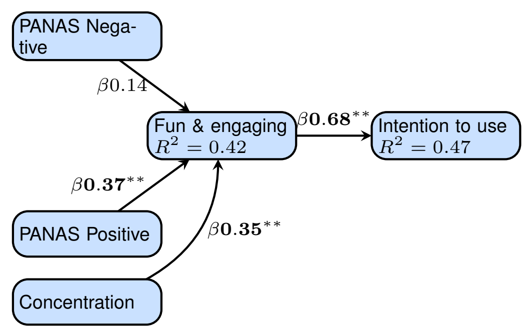


Figure 9: Research Model 1: PLS to predict the intention of use for all visitors. Independent variables: PANAS Negative (PN) and PANAS Positive (PP) are emotional factors; concentration is a cognitive factor; fun & engaging is a hedonic factor. ** = significant; $n = 67$

in terms of weight vectors.

PLS does not require a large sample size [76], and it is not a pre-requisite that the research models are based on comprehensive theories [77][78]. Still, a research model should have a theoretical foundation, although it might contain exploratory aspects.

We used statistical modelling with *smartPLS* version 2.0 M3 [79] to analyse the data and to compare the three models. TABLE V shows the composite reliability of the constructs and the factor loading for each item that need to meet the evaluation criteria for partial least square modelling [75]. We updated these results from Tjøstheim et al. [2] by increasing the number of respondents ($n = 67$). In TABLE V, note that C_2 has low factor loading. This is caused by winners answering differently from losers (winners tend not to read questions more than once).

We created a dependent variable *intention to use the eQuiz again*. The R^2 variable is a measure of the proportion a dependent variable is explained by the independent variables

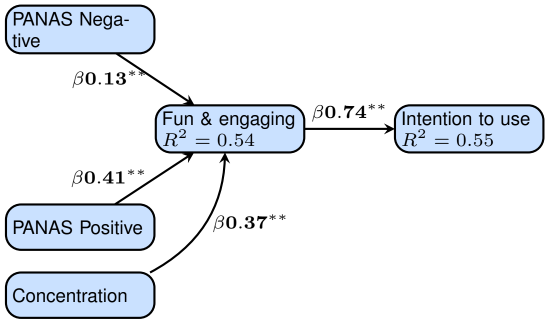


Figure 10: Research Model 2: PLS to predict the intention of use for losers. $** =$ significant; $n = 34$

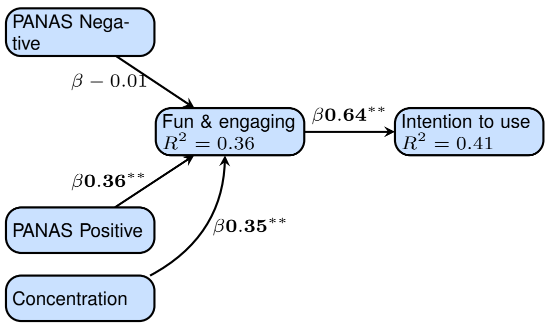


Figure 11: Research Model 3: PLS to predict the intention of use for winners. $** =$ significant; $n = 33$

in the model. For the winners and the losers of the game, the factors *fun* and *engaging* explain the new variable at 41% and 55%, respectively; combining them together the value is 47%. We updated the graphs to include data from more participants. Figures 9, 10, and 11 show graphical representations of these dependencies for all participants, losers, and winners.

For someone managing a science centre, it seems like a good choice to ask visitors if they enjoyed playing eQuiz. It is valuable to know the answer to this question, and it might give the science centre an indication of whether the visitors are interested in using the installation again.

B. Automating the Assessment Process

The installation *The Motorway of the Ocean* in the exhibition *At Sea* at the Norwegian Maritime Museum (NMM), here denoted as NMM-01, is a game that teaches players the roles of people employed in shipping and tasks related to shipping from the perspective of a ship owner.

In NMM-01, up to four players compete against each as ship owners. Each player controls one vessel. Through the course of the game, players make informed decisions as the ships travel across the ocean to its destination. These decisions can be the speed of the vessel, whether or not to take on extra cargo, bunker oil, deal with weather or pirates, and so on. Each player is placed behind a console where the player can control the game while the current progress for all ships is shown on a projection wall, visible for all.

For NMM-01, we used a similar setup as in the previous case, i.e., we used camera observation, observation by human observers using a simplified valence tracker, and questionnaires including PANAS, a questionnaire about emotions, and a questionnaire containing knowledge questions. The visitors answered this questionnaire both before and after the game was played. This questionnaire was developed to assess the impact of the game on visitor's knowledge.

We used Kinect II devices to create the video footage for the analysis. Additionally, we analysed whether emotions can be assessed using the Kinect-API that allows to retrieve the parameters *smile* and *engaged*. Currently, the software for this type of assessment is under development. The first tests show that smiles in the faces of the visitors can be recognised, but *engaged* currently only means that the visitor is facing the Kinect. Some technical issues with the API and a low number of test subjects has resulted in insufficient data so far.

The idea is to compare the results from the valence tracker, the face reader using the video-footage from the Kinect II, and the results from the Kinect-API. We noted that the nature of NMM-01 is not making people smile much; thus, the retrieved data are not sufficient to come to a conclusive answer.

Results from the questionnaires show that NMM-01 shows the highest values in terms of engagement, as can be seen for the value E in Figure 12.

C. The Influence of the Narrative to Experience

The installation *Solar Cell*, here denoted as NTM-01, is part of the exhibition *Energy Tivoli* at the Norwegian Museum for Science and Technology. The installation presents a wall where an atom and its electrons are drawn. The goal is to visualise how energy is created from colliding protons and election. The visitor throws balls representing photons at the four outer electrons on the wall during a given time. The original installation only instructed visitors to hit the outer electrons. For our experiment, the first half of the participants ($n = 39$) use the installation unchanged. The second half of the participants ($n = 36$) received more information. We extended the installation with an additional board explaining the role of electrons and photons before participants started throwing balls. We compute this as a change in VEI profile for the N-dimension from Level 2 to Level 3.

In our evaluation, we asked visitors to use NTM-01, either the original or modified version, and fill out the questionnaires as described for the other evaluations. The evaluation shows that the engagement increases slightly for the modified version, but this increase is not statistically significant with the number of participants that took part in the study. This result is expected since the changes to the narrative are small. We also noted that the participants were focused on throwing as many balls as possible within the time-limit of twenty seconds.

We observed that the unmodified NTM-01 shows a low score for the *intention to learn* (L_i); after the modification, this score increases by about one point on the Likert scale. We refer to Figure 12 explained in Section V-D for more details.

NTM-01 was included in our study because it is an example of a common type of installations in science centres that

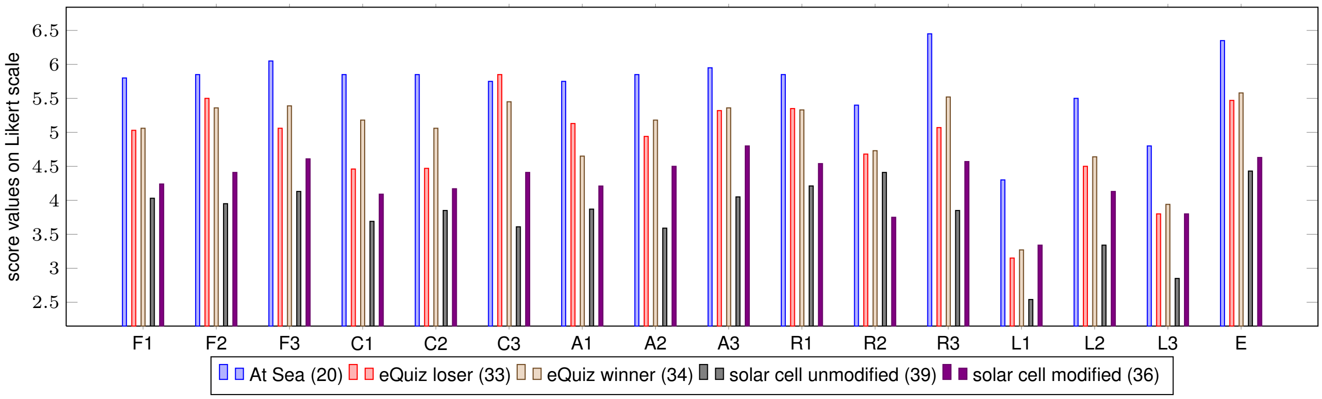


Figure 12: Response scores for five tests on a Likert scale for the independent variables for *fun* F_i and *concentrate* C_i and the dependent variables for *play again* A_i, *recommend* R_i, and *intention to learn* L_i. Each variable occurs in three different questions. We also show the results for the variable *engaging* E.

emphasise user's physical activity, the time use is rather short, and little written information is presented.

D. Comparing the installations

In Figure 12, we show the results from the questionnaires for the installations NMM-01, ENG-12 (separate results for winners and losers), and NTM-01 (modified and unmodified) in a graph for several dimensions of engagement. The graphs show the results for the independent variables *Fun* (F) and *Concentrate* (C), and the dependent variables *Play Again* (A), *Recommend* (R), *Intention to Learn* (L), and *Engaging* (E). The trend in these data is clear: NMM-01 has the highest engagement factor, followed by ENG-12, while both versions of NTM-01 score significantly lower.

For each of the variables F, C, A, R, and L, three different questions were asked; this was done to combat possible effects from the way the questions were presented. One question was asked for variable E. The formulation of the questions is shown in TABLE IV for ENG-12. For the installations NMM-01 and NTM-01, similar questions were asked with some modifications due to the nature of the installation.

As Figure 12 shows, there are some differences between the responses in one category, but most of these follow the same pattern; with the exception of R_2 score.

Our goal was to find a suitable proxy question for engagement. The analysis shows that the values for F_i are most structurally similar to the variable E. This result suggests that studies in science centres can use one of the questions about fun as a proxy for engagement instead of asking several questions to the visitors. It is straightforward to implement these shorter questionnaires in an application on a mobile device or another part of a system that facilitates studies in science centres and museums.

VI. Conclusion and Future Work

Science centres and museums are interested in having engaging exhibits to attract visitors. Our methodology to assess and analyse visitors' engagement can be a new instrument in finding these exhibits. Our assessments at three science centres used installations with different properties to find evidence that we can use measured values from installations, sensors, and cameras to estimate visitor engagement. We can use this evidence to reduce the size of questionnaires down to a few questions. Using the evidence and the short questionnaires gives us a good way to find and assess engagement.

This shows that our methodology can work in different kinds of science centres with different subject matter and ways of interacting with an exhibit. We have explored different methods for gathering data. We focused on emotions visible on the visitor's face combined with data from the installation, and short questionnaires, but other methods could also work such as skin conductivity or heart rate. There are few limits to data sources, but each source could add to complexity.

The science centres and those creating exhibits benefit from this research. They can use our tools in the design phase and in maintenance to find what engages visitors and how a change affects engagement. The tools described in the article, like the VEI profile, is actively used in the design phase of exhibitions. The results of the assessments have been integrated into an exhibition management system to give science centres a better means to perform visitor studies.

There are several research paths forward. Further refinement and validation of our work so far needs to be performed. This includes a comparison with results from researchers who use qualitative analysis methods, such as sense-making.

When extending our methodology to use more sensors and collecting all available data using the Internet of Things (IoT) [80], several challenges occur. The sensors of the IoT can produce large amounts of data that need to be analysed. This amount of data is much larger than being processed in traditional visitor studies. Methods described in the research field of *big data* [81] need to be applied, including statistical methods and machine learning. Collecting large amounts of data will result in challenges for the visitor's privacy that are beyond the usual privacy consent agreements in museums when visitor studies are performed. Analysing ample sensor data using big data technology could potentially identify visitors through their behaviour in situations when they should be able to experiment unobserved. Both legal and technological measures must be considered to find a balance between allencompassing visitor studies and privacy.

Finally, our current work does not take into account the visitor's identity type, as defined by Falk and Storksdieck [17], nor the visitor's expectation to a science centre visit. Our studies were performed with students who are a homogeneous group. But we suspect that the visitor's engagement will vary for the different identity types. Thus, an integration of the VEI profile with visitor type and the subjective context of a visitor needs to be considered.

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.

References

[1] W. Leister, I. Tjøstheim, G. Joryd, and T. Schulz, "Towards assessing visitor engagement in science centres and museums," in Proc. PESARO 2015, The Fifth Intl. Conf. on Performance, Safety, and Robustness in Complex Systems and Applications. IARIA, 2015, pp. 21–27.

[2] I. Tjøstheim, W. Leister, A. Larssen, and T. Schulz, "The role of emotion and enjoyment for QoE – a case study of a science centre installation," in Proc. QoMeX 2015, The 7th Intl. Workshop on Quality of Multimedia Experience, A. Skodras, Ed. IEEE, 2015, pp. 1–6.

[3] L. Rennie and D. Johnson, "Visitors perceptions of changes in their thinking about science and technology following a visit to science center," Visitor Studies, vol. 10, 2007, pp. 168–177.

[4] S. Nordal, "Science centres and museums as rooms of revelation (RoR) in science education," in Abstracts for seminar on Science Centre Research. skolelab.no, Sep. 2012. [Online]. Available: https://skolelab.no/innsendt/skolelab.no/kurs/711/2012-10-08_135131-Nordal_Abstract_DeWitt_seminar.pdf [Accessed: 1. June 2015].

[5] M. I. Dahl and D. Stuedahl, "Transforming children's participation and learning in museums: From singular dialogues to a multilayered explorative experience," in Proc. The Transformative Museum. Roskilde University, May 23-25 2012.

[6] P. M. F. Sandsmark, "Mellom opplevelse og læring – en undersøkelse av publikums bevegelse og interaksjon i utstillingen Skipet," bachelor's thesis, Høgskolen i Telemark, 2011, in Norwegian.

[7] 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.

[8] 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.

[9] 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].

[10] H. H. Higgins, "Museums of Natural History : Part I. Museum Visitors". D. Marples, 1884.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] P. M. McManus, "Museum and visitor studies today," Visitor Studies, vol. 8, no. 1, 1996, pp. 1–12.

[16] 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.

[17] 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.

[18] 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.

[19] J. H. Falk, J. Heimlich, and K. Bronnenkant, "Using identity-related 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.

[20] C. Barriault and D. Pearson, "Assessing Exhibits for Learning in Science Centers: A Practical Tool," Visitor Studies, vol. 13, 2010, pp. 90–106.

[21] H. H. Shettel, M. Butcher, T. S. Cotton, J. Northrup, and D. C. Slough, "Strategies for determining exhibit effectiveness," American Institutes for Research, Washington, DC, Tech. Rep. Report No. AIR E-95-4/68-FR, April 1968.

[22] M. B. Alt and K. M. Shaw, "Characteristics of ideal museum exhibits," British Journal of Psychology, vol. 75, no. 1, 1984, pp. 25–36.

[23] D. Spegel, "Expogon," 2014. [Online]. Available: www.expogon.org [Accessed: 1. June 2015].

[24] 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.

[25] S. Allen, "Designs for learning: Studying science museum exhibits that do more than entertain," Sci. Ed., vol. 88, no. S1, 2004, pp. S17–S33.

[26] 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.

[27] S. S. Yalowitz and K. Bronnenkant, "Timing and tracking: Unlocking visitor behavior," Visitor Studies, vol. 12, no. 1, 2009, pp. 47–64.

[28] J. Baur, Ed., "Museumsanalyse – Methoden und Konturen eines neuen Forschungsfeldes", ser. Kultur- und Museumsmanagement. Bielefeld: Transcript, 2010, in German.

[29] V. Kirchberg, "Besucherforschung in Museen: Evaluation von Ausstellungen," in Museumsanalyse: Methoden und Konturen eines neuen Forschungsfeldes, ser. Kultur- und Museumsmanagement, J. Baur, Ed. Bielefeld: Transcript, 2010, pp. 171–184, in German.

[30] S. Macdonald, Ed., "A Companion to Museum Studies", 1st ed. Wiley-Blackwell, Aug. 2010.

[31] C. Martindale, "Recent trends in the psychological study of aesthetics, creativity, and the arts," Empirical Studies of the Arts, vol. 25, no. 2, 2007, pp. 121–141.

[32] P. Pierroux and R. Steier, "Making it real: Transforming a university and museum research collaboration into a design product," in Design as Scholarship in the Learning Sciences, V. Svihla and R. Reeve, Eds. London: Routledge, 2016, ch. 9, pp. 115–126.

[33] G. Klein, B. Moon, and R. R. Hoffman, "Making sense of sense-making 1: Alternative perspectives," IEEE Intelligent Systems, vol. 21, no. 4, Jul. 2006, pp. 70–73.
[34] ——, "Making sense of sensemaking 2: A macrocognitive model," IEEE Intelligent Systems, vol. 21, no. 5, Sep. 2006, pp. 88–92.
[35] M. Tröndle, S. Greenwood, V. Kirchberg, and W. Tschacher, "An integrative and comprehensive methodology for studying aesthetic experience in the field: Merging movement tracking, physiology, and psychiological data," Environment and Behaviour, August 9 2012, pp. 1–34.
[36] M. Oppermann, "Triangulation – a methodological discussion," International Journal of Tourism Research, vol. 2, no. 2, 2000, pp. 141–146.
[37] W. Tschacher, S. Greenwood, V. Kirchberg, S. Wintzerith, K. Van Den Berg, and M. Tröndle, "Physiological correlates of aesthetic perception of artworks in a museum." Psychology of Aesthetics Creativity and the Arts, vol. 6, no. 1, 2012, pp. 96–103.
[38] M. Ghijsen, "Facial expression analysis for human computer interaction," Ph.D. dissertation, University of Twente, 2004.
[39] V. Terzis, C. N. Moridis, and A. A. Economides, "Measuring instant emotions based on facial expressions during computer-based assessment," Personal and Ubiquitous Computing, 2011, pp. 1–10.
[40] J. A. Russell, "A circumplex model of affect," Journal of Personality and Social Psychology, vol. 39, 1980, pp. 1161–1178.
[41] H. L. O'Brien, "The influence of hedonic and utilitarian motivations on user engagement: The case of online shopping experiences," Interact. Comput., vol. 22, no. 5, Sep. 2010, pp. 344–352.
[42] K. S. Judge, C. J. Camp, and S. Orsulic-Jeras, "Use of montessori-based activities for clients with dementia in adult day care: Effects on engagement," American Journal of Alzheimers Disease and Other Dementias, vol. 15, no. 1, 2000, pp. 42–46.
[43] J. Griffin, "Learning science through practical experiences in museums," International Journal of Science Education, vol. 20, no. 6, 1998, pp. 655–663.
[44] D. Griffin and L. Paroissien, Eds., "Understanding Museums – Museums, education and visitor experience". National Museum of Australia, 2011. [Online]. Available: http://nma.gov.au/research/understanding-museums/Museums_education.html [Accessed: 1. June 2015].
[45] F. H. Wilhelm and P. Grossmann, "Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment," Biological Psychology, vol. 84, 2010, pp. 552–569.
[46] L. E. Nacke and C. A. Lindley, "Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience," CoRR, 2010. [Online]. Available: http://arxiv.org/abs/1004.0248 [Accessed: 1. June 2015].
[47] A. G. Money and H. Agius, "Analysing user physiological responses for affective video summarisation," Displays, vol. 30, no. 2, Apr. 2009, pp. 59–70.
[48] R. Picard, "Affective computing," MIT Media Laboratory, Perceptual Computing Section, Tech. Rep. 321, Nov. 1997.
[49] M. E. Hoque, J. Hernandez, W. Drevo, and R. W. Picard, "Mood meter: Counting smiles in the wild," in Proc. UbiComp'12. New York, NY, USA: ACM, 2012.
[50] M. Ben Ammar, M. Neji, A. Alimi, and G. Gouardères, "The Affective Tutoring System," Expert Systems with Applications, vol. 37, no. 4, 2010, pp. 3013–3023.
[51] M. Brand, F. Klompmaker, P. Schleining, and F. Weiß, "Automatische Emotionserkennung – Technologien, Deutung und Anwendungen." Informatik Spektrum, vol. 35, no. 6, 2012, pp. 424–432, in German.
[52] H. J. Witchel, C. E. I. Westling, J. Tee, A. Healy, R. Needham, and N. Chockalingam, "What does not happen: quantifying embodied engagement using nimi and self-adaptors," Participations, vol. 11, no. 1, May 2014, pp. 304–331.
[53] H. J. Witchel, C. E. I. Westling, J. Tee, R. Needham, A. Healy, and N. Chockalingam, "A time series feature of variability to detect two types of boredom from motion capture of the head and shoulders," in Proceedings of the 2014 European Conference on Cognitive Ergonomics, ser. ECCE '14. New York, NY, USA: ACM, 2014, pp. 37:1–37:5.
[54] K. Huotari and J. Hamari, "Defining gamification: a service marketing perspective," in International Conference on Media of the Future, Academic MindTrek '12, Tampere, Finland, October 3-5, 2012, A. Lugmayr, Ed. ACM, 2012, pp. 17–22.
[55] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: defining "gamification"," in Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, ser. MindTrek '11. New York, NY, USA: ACM, 2011, pp. 9–15.
[56] K. M. Kapp, "The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education", 1st ed. Pfeiffer & Company, 2012.
[57] D. Dixon, "Player types and gamification," Proceedings of the CHI 2011 Workshop on Gamification, 2011.
[58] R. A. Bartle, "Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs," Journal of MUD Research, 1996.
[59] A. Marczewski, "User types," in Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Design. CreateSpace Indep. Publ. Platform, 2015, pp. 65–80.
[60] N. Legault, "Gamification techniques: How to apply them to e-learning," E-Learning Heroes, March 2015. [Online]. Available: https://community.articulate.com/articles/gamification-techniques-how-to-apply-them-to-e-learning [Accessed: 1. June 2015].
[61] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work? – a literature review of empirical studies on gamification," in System Sciences (HICSS), 2014 47th Hawaii International Conference on, Jan 2014, pp. 3025–3034.
[62] P. Weiser, D. Bucher, F. Cellina, and V. De Luca, "A Taxonomy of Motivational Affordances for Meaningful Gamified and Persuasive Technologies," in Proc. 3rd International Conference on ICT for Sustainability (ICT4S), ser. Advances in Computer Science Research, vol. 22. Paris: Atlantis Press, 2015, pp. 271–280.
[63] C. M. Bishop, "Pattern Recognition and Machine Learning (Information Science and Statistics)". Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
[64] W. Leister, "Towards a generic assessment framework – ideas and cases," Norsk Regnesentral, NR Notat DART/02/2015, 2015.
[65] W. Leister and I. Tjøstheim, "Concepts for User Experience Research," Norsk Regnesentral, NR Note DART/14/2012, 2012.
[66] M. Cavazza and D. Pizzi, "Narratology for interactive storytelling: A critical introduction," in Technologies for Interactive Digital Storytelling and Entertainment: Proc. Third International Conference, TIDSE 2006, Darmstadt, Germany, December 4-6, 2006, S. Göbel, R. Malkewitz, and I. Iurgel, Eds. Berlin, Heidelberg: Springer, 2006, pp. 72–83.
[67] J. Hamari and V. Eranti, "Framework for designing and evaluating game achievements," in Proc. 2011 DiGRA International Conference: Think Design Play. DiGRA/Utrecht School of the Arts, January 2011.
[68] 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.

[69] B. Zaman and T. Shrimpton-Smith, “The facereader: Measuring instant fun of use,” in Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles, ser. NordiCHI ’06. New York, NY, USA: ACM, 2006, pp. 457–460.

[70] 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.

[71] 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.

[72] E. Jaffe, “The psychological study of smiling,” APS Observer, vol. 23, no. 10, 2010.

[73] M. G. Frank and P. Ekman, “Physiologic effects of the smile,” Directions in Psychiatry, vol. 16, no. 25, December 1996, pp. 1–8.

[74] H. A. Landsberger, “Hawthorne revisited: management and the worker : its critics, and developments in human relations in industry”. Cornell University, 1958.

[75] W. W. Chin, “The partial least squares approach to structural equation modeling,” in Modern Methods for Business Research, G. A. Marcoulides, Ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1998, pp. 294–336.

[76] C. Fornell and D. F. Larcker, “Evaluating structural equation models with unobservable variables and measurement error: Algebra and statistics,” Journal of marketing research, vol. 18, no. 3, 1981, pp. 382–388.

[77] D. Barclay, C. Higgins, and R. Thompson, “The partial least square (PLS) approach to casual modeling: Personal computer adoption and use as an illustration,” Technology, vol. 2, no. 2, 1995, pp. 285–309.

[78] W. W. Chin, B. L. Marcolin, and P. R. Newsted, “A partial least squares latent variable modeling approach for measuring interaction effects: Results from a monte carlo simulation study and an electronic-mail emotion/adoption study,” Info. Sys. Research, vol. 14, no. 2, Jun. 2003, pp. 189–217.

[79] C. M. Ringle, S. Wende, and A. Will, “SmartPLS 2.0 (M3),” 2005. [Online]. Available: www.smartpls.de [Accessed: 24. Feb 2015].

[80] K. Ashton, “That ’Internet of Things’ Thing,” RFID Journal, 22 July 2009.

[81] J. S. Ward and A. Barker, “Undefined by data: A survey of big data definitions,” CoRR, vol. abs/1309.5821, 2013.