Towards Reliable Accessibility Assessments of Science Center Exhibits

Till Halbach[(✉)] and Ingvar Tjøstheim

Norwegian Computing Center, Oslo, Norway
{till.halbach,ingvar.tjostheim}@nr.no

Abstract. A methodology for assessing the degree of accessibility of museum and science center exhibits is proposed in this work. It consists of an exhibit-centric framework with a corresponding score for the accessibility indicators vision, hearing, mobility, motor, and cognition. The scores are achieved by combining the results of an observational study with a supplementary expert evaluation for high reliability. The methodology was successfully applied with pupils in an informal-learning context at Oslo Science Center. We believe it has relevance for and can be applied to other areas such as self-service machines and interactive technical artifacts in general due to its generic nature.

Keywords: Accessible interactives · A11Y · Universal design · UD · Digital inclusion · Museum · Exhibition · Installation · Visitor experience

1 Introduction

In how far are exhibits at museums and science centers universally designed? Are there the same opportunities for pleasing visitor experiences for all, regardless of ability or disability? If not, what kind of barriers do visitors with disabilities encounter? These are some of the questions addressed in this study, for which we cooperated with Oslo Science Center. These centers target a very broad audience, but in this study we have focused on how pupils experience and use interactive and electronic artifacts.

According to statistics from 108 Norwegian museums for 2017 and 2018, 73% of the museums have employed accessibility measures in terms of braille labeling, use of sign language, large print, and easy-to-read information [1]. We believe that these numbers are biased because they are based on institution self-reports, and they lack the proper documentation and description of methodology. Also, the list of criteria is very limited and does not cover important aspects like multimodality, video captions, color contrasts, etc. There is anyhow a potential for improvement, given that 15-20% of the Norwegian population have some form of disability [2].

The work is organized as follows: We present the assessment framework and corresponding measurement methods first. Then it is demonstrated how exhibits at Oslo Science Center can be assessed with school-children as visitors. Finally, we present our conclusions and outlook for future research.

2   Related Work

One of the earliest publications regarding museums and accessibility dates back to 1977 and originates from The Smithsonian Institution [3]. The Smithsonian has later published extensive and detailed guidelines linked to their dedicated Accessibility Program [4]. Related to this are the recommendations for universal design of exhibitions in science museums, which is the work of a number of experts in the field with the lead of the Museum of Science in Boston [5]. Included in this document is also a checklist, which can be used to get a quick impression for how inclusive an exhibition is. An earlier research report from the Museum of Science studied not only the visitor experiences of persons with disabilities but also the relationship of universal design and informal learning contexts in general [6]. The author argues that some accessibility measures for one particular group may counteract the measures for other visitor groups with different needs, but most of the accessibility measures also benefit persons without disabilities.

In general, accessibility measures in museums are necessary but not sufficient for learning and interest [7]. A strategy to achieve both is the application of recommendations from Universal Design for Learning, or short UDL [8]. Basic elements here are the use of multiple modalities, specific measures for the creation of engagement and motivation, and giving visitors several options for interaction and ways of expressing themselves. Rappolt-Schlichtmann argues that the UDL recommendations not only can give visitors with disabilities access to exhibits but may also lead to increased motivation and improved learning for all. Many factors for learning in science museums are, however, not sufficiently understood [9], such as effect of the surroundings and context. According to the author, measures could be counteracting if used without reflection about their impact, and there are many potential pitfalls: For instance, confusion caused by too many selection options and choices, distraction due to many interaction possibilities, disruptions through simultaneous interaction mechanisms, and failing to grasp the main learning objective due to multiple content alternatives. The author concludes that there are no generic solutions, and that the key for good learning experiences is the combination of user-centric design and universal-design measures with a special focus on cognitive accessibility.

3   Assessment Framework

It is common to categorize disabilities into three main areas: *Sensor*, *motor*, and *cognition* [10]. However, it makes sense to distinguish between *hearing* and *vision* as the most important sensors. This is also done by the WHO. Next, *motor* is a broad term which does not adequately mirror the difference between using one's legs, in our study referred to as *mobility*, and the use of hands and arms to interact with something, which is the meaning used here. In addition, we suggest a category for voice control. This gives the six areas *vision* (V), *hearing* (H), *mobility* (M), *motor* (MT), *voice* (VC), as well as *cognition* (C), here called accessibility indicators. *Cognition* compounds a variety of processes, including orientation, language, reasoning, memory, concentration, coordination, learning and engagement.

The guidelines on universal design of self-service machines [11], by Norway's Agency for Public Management and eGovernment (Difi), are also relevant for interactive exhibits. The agency recommend the following four main areas for development and assessment: *Finding the machine*, *getting there*, *surroundings and area of use*, and *use of the machine*. While this is a useful approach, it lacks two important topics, which are addressed in the WCAG 2.1 standard [12]: *perceivable* and *understandable*. Combined with Difi's areas, this leads to the topics *getting to/from*, *perceive*, *control*, and *understand*. *Findability* is then part of *perceive*, *use* belongs to *control*, and *surroundings and area of use* relates to *getting to/from*. WCAG's *robust* principle is no longer an area of its own but included with the other four; see Fig. 1. The inner circle is associated with the visitor's movement to a particular exhibit or installation. Within interaction distance, the visitor's actions are dominated by a combination of perceiving, controlling, and understanding (outer circle). The disability indicators can now be mapped to exhibit areas: *Mobility* relates to *getting to/from*, *vision* and *hearing* to *perceive*, *motor* and *voice* to *control*, and *cognition* to *understand*.

To quantify each indicator, we propose a grading system with four levels/points. This number is a compromise considering that the scale should have as few levels as possible, be easy to understand, and allow for quantifying different degrees of (in-) accessibility. The levels are:

- 1 point: Absolute barrier(s). A true "showstopper", no work-around.
- 2 points: Significant barrier(s). Increased use of time, high risk for making mistakes, as well as significantly reduced user experience.
- 3 points: Minor barrier(s). Somewhat reduced user experience.
- 4 points: No barrier. Normal user experience.

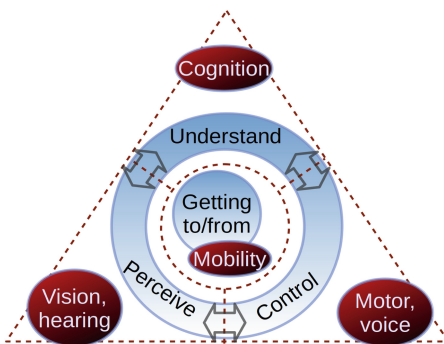In the following, it is discussed how the rating can be compiled.

Fig. 1. Mapping of accessibility indicators to exhibit areas (left), and the exhibit "Running track" (right)

4   Methodology

In accessibility and usability research, it is the user or group of users who have the final verdict for how accessible or usable a given solution is [13]. Therefore, trials and studies involving a wide range of users with different backgrounds are mandatory for correct results [10]. Unfortunately, this is often difficult to achieve as recruiting users to form a representative population is not a straightforward process. Moreover, user trials may give inconclusive or contradictory results for different user groups [7], and conducting them can be quite time-exhaustive and expensive [14].

Evaluations carried out by experts are usually cheaper and more efficient in terms of the number of accessibility issues found per invested time [14]. Also, trained experts typically have the knowledge needed to bridge the gap between opposite user groups. On the other hand, experts might not get the same results as studies with actual users, and different experts do not necessarily agree in their conclusions [15]. A combination of user assessment with expert assessment can therefore be the right strategy to reduce the weaknesses with each method. The findings from the user trials should guide the experts in their own assessment, minimizing the potential for diverging results and disagreements. Moreover, the experts can interpret inconclusive user trial data, use their knowledge to fill potential gaps, and collect more data points where this is possible.

To determine the appropriate indicator level in the assessment framework, all data points in terms of observation notes, quotes, questionnaire answers, camera recordings and others are mapped to one of the accessibility indicators. In case of missing data points, experts can conduct empathetic walk-throughs and/or sessions with simulated impairments and a particular accessibility area in mind and note down any challenges and barriers. With the given definition of accessibility indicator levels, the resulting (final) dataset is then used to decide how many points an exhibit should be given in a particular area. It is stressed that each exhibit typically relates to multiple indicators.

For instance, the position and size of a button, as well as how easy it can be pushed, is associated with the indicator *motor*. Text on the button, and in how far it is lightened up, is mapped to *vision*. How well it can be understood what it controls is related to *cognition*. A button is usually not linked to *mobility*, *voice* or *hearing* in any way (if there is no auditory feedback upon a button push) and has hence no influence on the choice of levels in these areas.

5   Case Study

We have carried out a first step in verifying the assessment framework and proposed method through a case study that included the assessment of 15 exhibits at Oslo Science Center, undertaken in Autumn 2018. Oslo Science Center is primarily targeting school classes and family visits, meaning that most visitor experiences happen in small groups, and that there are no real restrictions to the visitor age. Pupils with the ability to read and general knowledge in math and science, however, may benefit the most from a visit.

We recruited 34 informants in total, 25 pupils and nine adults, distributed over two families and three school classes. The pupils were between nine and 12 years old, and almost two out of three were male. Some of them had visited the Science Center

previously, a few even several times. Three had low vision, 23 had hearing disabilities (both cochlear implant and other hearing aids), and one had cerebral parese. We assume that many also had hidden impairments like dyslexia and learning difficulties. Among the adults, there were two parents, three teachers, one assistant, and three interpreters.

The pupils could choose their own route around the venue either alone or in small groups of up to six. They were typically followed by teachers and/or interpreters, as well as observing researchers. The observers were passive for most of the time but sometimes had to intervene the situation with explanations and demonstrations, and they also asked questions about the pupils' experience during and after a session, which lasted 45 to 60 min. Thus, when the collection of user data was finished, there was a multitude of data points in terms of observation notes, quotes, questionnaire answers, and camera recordings.

The objective of the subsequent expert assessment was, as already mentioned, to add data points for missing accessibility areas. No exhibit had voice control, so there were no *voice*-related challenges. Next, we had a lot of user data regarding *vision* and *hearing* but few data points related to *mobility* and *motor*. Our expert assessment focused therefore on both areas and was guided by technical report ISO TR 22411:2008, which – while not being an international standard – provided useful recommendations for the optimum height of control elements, minimum width for passages and other areas, maximum reach to screens and controls, minimum audio volume, and more. When it comes to *cognition*, it is stressed that it was outside the scope of this project to thoroughly collect data regarding whether or not the pupils had understood the given phenomena, and what they might have learned after their visit. We did, however, have a number of indications from the user trials regarding the pupils' comprehension and engagement. The expert evaluation thus concentrated on a limited set of cognitive aspects, such as language- and reading-related issues, dyslexia, as well as basic reasoning.

6 Example Assessment

We illustrate the use of the assessment framework and proposed method by means of a concrete example, the exhibit "Running track" at Oslo Science Center, depicted in Fig. 1 (right side). In the following, the various assets and aspects of an exhibit are coded with appropriate accessibility indicators to clarify their contribution to the given indicator level (V, H, M, MT, VC, C).

The exhibit's learning purpose is to teach its users about the relation of distance, velocity, and time. Two participants compete in running 10 m beside each other and as fast as they can on tartan tracks. Prior to this, a countdown must be started by pushing a button, counting both on a screen and on a loudspeaker (3-2-1-"shot"). The participants run straight ahead until they pass the finish line (and are eventually stopped by a mattress against the wall), upon which their time is taken, and a slow-motion video of the participants' performance is shown together with their result (on the finish screen). It is possible to run alone, too, but more fun doing so pair-wise. On the days of the trials, the starting screen was out of order (V), leaving only the auditive countdown for its users, and sometimes the finish line sensors would not capture the correct running time (C), confusing the users with irrelevant numbers.

The corresponding label associated with the exhibit is found on a nearby wall at a readability-friendly height (V) and uniquely identifies (C) the exhibit, it gives instructions (C) and tries to inspire the visitor by the use of illustrations and engaging questions (C), and it explains the exhibit's learning purpose (C). It contains text in both Norwegian and English - a plus for non-native speakers (C) - with short and easily comprehensible sentences (C). (Labels at other exhibits sometimes contain technical terms which are too difficult for the majority of pupils, let alone those with reading/learning difficulties.) The English text is solely set in italics, which should be avoided due to poor readability (V). Text and illustrations have sufficient contrast (V), but an x-height of 3 mm means that the font size is too small (V). We noticed that pupils with low vision had to get really close (10–20 cm) to the labels to be able to read them. (Labels at other exhibits sometimes have a glossy surface with many undesirable reflections (V).) There are no auditive labels or provision of other modalities for entirely blind people (V).

The exhibit is located in the center of the exhibition, well illuminated, and very easy to find/get an eye on (V). There is plenty of space around and no obstacles except for a small 3-cm edge around the tartan floor, which can be problematic for wheelchairs (M). The luminescent (V) countdown button is easy to reach, easy to target, and easy to push (MT). It is also possible to adjust the starting blocks, but this requires some finger strength (MT).

Both screens and the loudspeakers are placed in a readability- and hearing-friendly height but only mounted in one position on the left wall (V, H). The pupils with hearing impairments in our trials had to turn their heads away from the tracks and towards the wall in order to capture the signal. In addition, the audio is quite low (roughly 48 dB near the starting block farest away), challenging those with hearing aids in our trials (H). Both screens have sufficient brightness and contrast (V), but the font size is only average (V), meaning that some pupils with low vision had to stand quite close to it to be able to read, and sometimes this conflicted with those running. Had the starting screen worked, it would have shown some instructions ("get ready for countdown") which are not given by the loudspeaker (V). There is no result service for entirely blind users (V).
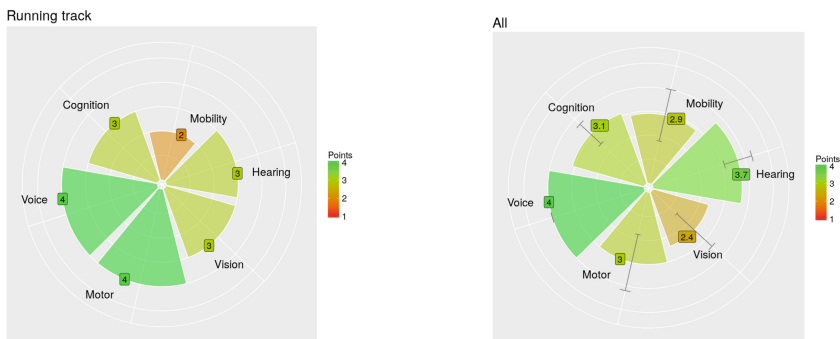

Fig. 2. Example assessment of the exhibit "Running track" (left), and mean and standard deviation of the accessibility indicators for all evaluated exhibits (right)

According to notes from the observers, the vast majority of pupils understood the exhibit's concept right away (C). A few did not notice the starting button, most likely because it is not prominent enough (V, C), and almost nobody understood where the finish line was (C); they all run until they hit the mattress. The result screen shows the time and average and maximum speed of both participants, but does not inform about which track the numbers relate to, and who has won (C). In our trials, this lead to situations where pupils argued about the "ownership" of numbers, and sometimes they were unable to say who had won, because they were not capable of comparing two given numbers. Also, the correct understanding of the unit "km/h" cannot be expected by pupils in the lower ages (C).

All in all, this was a very engaging exhibit and it meant a lot of fun to all who tried, as we could easily tell from the pupils' reactions and from how much time they spend here. Even wheelchair users could use it (if the change in floor level is reduced), and with some minor modifications (an auditory result service) even blind visitors. It was interesting to see that even though many with hearing impairments encountered problems with the countdown, they were able to cope with this by helping each other, by counting themselves and giving signs to each other for when to start. The learning purpose of this exhibit is likely a bit overrated; its inspiration factor is quite high, though.

Figure 2 (left) summarizes our assessment. There are no *voice*- or *motor*-related barriers in this exhibit, so both indicators yield a maximum point score (4 points). The problems encountered on the areas *vision*, *hearing*, and *cognition* are all rated as minor as specified in the definition of indicator levels above, so 3 points here. The edge on the floor is viewed as severe for those with *mobility* issues, and it is hence given only 2 points. There are no absolute barriers with this exhibit, so no indicators are on the minimum point score.

We have followed the same procedure for 15 exhibits in total and, based on this, calculated the average levels including the standard deviation (avg./std.dev.) for each indicator. The resulting numbers indicate where there is most potential for improvement with the exhibition, and where accessibility measures vary the most. As can be seen in Fig. 2 (right side), the fewest challenges lie in the areas *voice* (4.0/0.0) and *hearing* (3.7/0.6). No exhibit is voice-controlled yet, and with many, hearing is not required in order to use them. (However, some of the other exhibits show phenomena related to sound waves.) We have found minor hinders in the areas m*obility* (2.9/1.1), *motor* (3.0/1.2), and *cognition* (3.1/0.6). The high standard deviation for *mobility* and *motor* can be explained by the fact that there are exhibits which have inherent absolute barriers and only achieve 1 point. Obviously, v*ision* (2.4/1.0) is the area with the least degree of accessibility. This is because most of the exhibits require vision to some extent, and virtually all have indisputable barriers for entirely blind visitors.

7   Limitations

Both framework and method have been in use in a single study only. In our trials, we have used a qualitative approach with relatively few informants (34) and few (15) exhibits. In particular, our observations would have benefit from informants with mobility and motor impairments, and also from pupils with different ages and severe cognitive impairments. However, we believe that we have minimized the problem of a non-representative population by the aforementioned expert assessment.

We are aware of the fact that all observers and expert opinions are to some extent subjective, and we have tried to counterbalance this effect by summarizing all observations and impressions across all researchers for each exhibit asset and exhibit. It would nevertheless have been an advantage to have multiple experts evaluate the exhibits in the last phase, and to have a higher number of exhibits for the statistical analysis. In this project, with the given project budget, though, this was not feasible. On the other hand, more experts would not necessarily have led to different results as most of the expert assessment is based on rather objective measures, such as the width of a passage, the height of buttons, a screen's font sizes, the volume settings of loudspeakers, and so on.

8   Conclusion and Outlook

In this work, a novel framework and methodology for assessing the degree of accessibility (and hence universal design) of museum and science center exhibits has been proposed. The framework is exhibit-centric and consists of six indicators and a score from one to four (where four is best). The indicators can be bar-plotted graphically for easy comprehension. The assessment method is characterized by a user study followed by a supplementary expert evaluation. The entire approach is rather generic and can thus be applied to other areas, too, such as self-service machines and interactive technical artifacts in general. Framework and method have been successfully tested by the assessment of 15 exhibits at Oslo Science Center.

It can be concluded that the approach is well suited to measure and quantify an exhibit's degree of accessibility and uncover its strengths and weaknesses in a reliable manner. The approach is demanding with regard to time consumption and costs, though. For future research, we suggest to compare the assessments from user trials and expert trials to determine if experts (and if so, how many) can reliably replace users.

An important observation from our trials is that both engagement and social aspects can counterbalance the downsides with barriers and hinders in informal learning. Engaged visitors usually do not give up so quickly when they meet barriers, and peer visitors can help working around hinders and also contribute to engagement. On a final note, a high number of exhibits in an exhibition is an accessibility measure in itself, as visitors who encounter barriers often will leave for other exhibits with more pleasing experiences. This argument should not, though, serve as a resting pillow for not addressing existing accessibility issues with exhibits, as everybody should be given the same opportunity to experience all phenomena in a science center, as called for by the UN's Convention on the Rights of Persons with Disabilities [16].

References

1. Kulturrådet. Statistikk for museum 2017 (2017)
2. Otnes, B., Ramm, J.: SSBs Levekårsundersøkelse. Statistisk sentralbyrå (2016)
3. Snider, HW.: Museums and Handicapped Students: Guidelines for Educators. Smithsonian Institution (1977)
4. Majewski, J.: Smithsonian Guidelines for Accessible Exhibition Design. Smithsonian (2011)
5. Reich, C.: Universal Design Guidelines for Public Programs in Science Museums. Boston Museum of Science & NISE Network (2008)
6. Reich, C.A.: Universal Design of Interactives for Museum Exhibitions. Museum of Science, Boston (2005)
7. Rappolt-Schlichtmann, G., Daley, S.G.: Providing access to engagement in learning: the potential of universal design for learning in museum design. Curator Mus. J. **56**(3), 307–321 (2013)
8. CAST. Universal design for learning guidelines. National Center for Universal Design for Learning (2011)
9. Allen, S.: Designs for learning: studying science museum exhibits that do more than entertain. Sci. Educ. **88**, S17–S33 (2004)
10. Fuglerud, K.S.: Inclusive Design of ICT: The Challenge of Diversity. University of Oslo, Faculty of Humanities (2014)
11. Agency for Public Management and eGovernment (Difi) Krav til automater|Universell utforming. In: uu.difi.no. https://uu.difi.no/krav-og-regelverk/automater/krav-til-automater. Accessed 2 Apr 2019
12. W3C. Web Content Accessibility Guidelines (WCAG) (2018)
13. Norman, D.: The Design of Everyday Things: Revised and Expanded. Basic books, New York (2013)
14. Røssvoll, T.H., Fuglerud, K.S.: Best practice for efficient development of inclusive ICT. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2013. LNCS, vol. 8009, pp. 97–106. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39188-0_11
15. Brajnik, G., Yesilada, Y., Harper, S.: Is accessibility conformance an elusive property? a study of validity and reliability of WCAG 2.0. ACM Trans. Access Comput. **4**(2), 8 (2012)
16. United Nations. Convention on the Rights of Persons with Disabilities (2006)