International Research Journal of Computer Science (IRJCS)
Issue 01, Volume 3 (January 2016)
ISSN: 2393-9842
www.irjcs.com

Game Approachability: Remodelling testing of computer Games with the GAID model

Jonas Molin
Department of informatics
Umeå University, Sweden

Charlotte Wiberg
Department of Informatics
Umeå University, Sweden

Abstract-- This paper reports on a large study conducted with the purpose of adding to our understanding of game approachability and, on the basis of this broadened understanding refine methods for evaluating this quality in games. With better tools for evaluation, better results from evaluations can be obtained which will result in refined game design. This is important in general, of course. However, with the new situation in game industry, where game designers create games for completely other types of users than themselves, which is a shift from designing for hardcore gamers, as themselves, to casual gamers, with little or no earlier skills in gaming, evaluations play a central part in the game design process. The paper reports on the results of the study in the form of a new model for game approachability evaluation – GAID. The model, a flow chart, is thoroughly presented and validated with examples from the study. Finally, conclusions are made on the potential of GAID, both in future research as well as its usefulness in game design industry.

Keywords— Usability; Game design; Game Approachability; Evaluation Methods; User Experience; Qualitative study

I. INTRODUCTION

With new types of media and new target groups as users, the need of new and refined design- and evaluation methods evolves (c.f. [1]). As the game industry grows the need for more well established research becomes more real [2]. Where the games of old days were designed by gamers for gamers the game industry of today have to cater to a much more unforgiving and novice audience than that of yesterday the need to make flawless game accessibility becomes a higher priority [3]. As parts of the design of these, the flawless games that are demanded by the consumers, the design/development process of use case studies is perhaps the one part that is totally sacred to the production of a good game. Use of case studies demand huge resource commitment by the developers in the form of experts in game design or usability studies (c.f. [4], [5], [6]). Much of the pitfalls of bad game design can be avoided by lists and heuristics [7] and thereby decrees the error rate during production, but the opinion of the user's studies will never be uninteresting. As long as game developers are producing original content (something that is expected from good games) there is a risk and situations will manifest that is not covered in heuristics. By making the tools of observational studies accessible to the less experienced in the ways of game design the industry will be able to use bigger samples in their use case studies whiteout having to dedicate entire armies of experts to the task. Supported by these reasons it is of interest to the HCI community to develop better, more simple tools to identify the flaws of games in the making.

[1] In this paper an attempt is made to improve on the heuristic model presented by Desurvire and Wiberg, first published in their paper published at the CHI 2008 conference [8]. The goal is to create a tool to help identify the types of problems that are found in the following articles on GAP [9], [10]. The improvement that is presented in this work is a flowchart that uses material from GAP. This will simplify the identification of limitations in game designs, using the material collected by H.Desurvire, and C. Wiberg [9], [10]. The benefit that the implementation of a flowchart hopes to bring to the research and design society is to make the GAP tool more accessible to less informed users. The approach taken in this work, to strengthen the GAP tool was inspired by the solution presented by Barendregt, Bekker and Speerstra in Empirical evaluation of usability and fun in computer games for children [11], who made a flowchart to help identify the type of problem that could be found in connection to children playing computer games. The function of such a chart in comparison to a list is to make it more available to personal less experienced in game design and the rules of good game design [12]. The flowchart was tested on an expert to verify the issues being used in it and on a usability test material. It was found that the Game Approachability Issue Definition (GAID) flowchart is able to identify, approachability issues, and thereby did improve on GAP as intended.

II. THE BASICS – USABILITY, ACCESSIBILITY, APPROACHABILITY AND PLAYABILITY

Usability, playability, approachability and accessibility are key concepts to the discussion taking place in this report and therefore deserve an explanation. They are the factors that are measured by GAP, PLAY and HEP, and later on also in GAID.

A. Usability

Usability is a factor in the discussion, but not one that will be focused on as it is a to general one. It is enough to say that usability is the factor by which the intended use of an artefact is performed. Usability therefore contains all of the above factors used to measure the success of intended use of an artefact, in this case a game [4], [5], [6], [13].

By this definition usability can measure any artefact, also a game, but as will be explained below, more specific –abilities are used in the field of game design focused HCI.

B. Accessibility

Accessibility is the factor by which the user of an artefact can go from no knowledge of the artefact to adequate knowledge to be able to use the artefact to satisfaction [14]. Accessibility is something that should be high in both games as in general software. The difference worth to mention between general software and games is that while utility software might demand a quick learning curve i.e. a high accessibility to reach expert knowledge with minimal effort, entertainment software desires a quick learning of initial knowledge but a slower road to expert knowledge which might lead to boredom as the challenge of solving the task is the innate goal of a game. In the words of Malone concerning games, "easy to learn, difficult to master" [12].

C. Approachability

Approachability in the context that is being discussed in this paper is about how easy it is for a novice player to be able to start playing the game without facing setbacks that will discourage the player from continuing [8], [9], [10]. As with the other abilities described in this section approachability shares many aspects with the rest. The qualities that sets approachability aside is its focus on the transition from, as might be learned from the name, novice user of a system (game in this case) to informed user. This is of critical value to the gaming industries in comparison to the rest of the commercial software market due to the need to satisfy the user that entertainment software is charged with. It is not that the other software are not made to satisfy the user, but in the case with entertainment software there is a certain expectation from the user to be entertained which is not present in the case of, for example, a word processing software were it more likely is that the software shall help the user to produce nice looking and correct text. By this argument the entertainment software industry has to be more compliant to the needs of their novice users than the designers of utility software who have to put a bigger focus on the goal of their product. The difference is that in the case of entertainment software, *enjoyment is* the goal. The use of the product is the goal of the product. Unless the user likes the process of beating the game, the way that takes them to that goal, the game designer has failed with the product. Approachability measures that the user can start to enjoy the product as soon as possible and thereby a form of usability of the game. The GAP list mentioned above focuses on isolating the parts that makes up for good approachability in games. If these parts can be identified and pretested for weaknesses a better product can be put to usability testing instead of having to plague the users with beta stages and high patching ratio in the early stages of the games launch to the market.

D. Playability

The playability of a game is the user's possibility to act in the game [15]. If a game is well designed the actual embodiment of the user's intention to manipulate the game in a certain way is the Playability of the game. This should however not be confused with the similar ability of usability often referred to in other software literature. The main difference between the two, similar as they may be, is that in the case of usability the designer strives to make all intended (and in some cases unintended interaction) goals as easy to reach as possible. In the case whit playability this is not the case as it would completely ruin the challenge of the game, and thereby the cardinal reason to play games [16]. The playability is the way to further the intention of action, not as much as the intention of reaching the goal of the action. A match of Chess, a game known to everyone, can serve to show the difference between the two abilities. From a usability point of view we want to beat the opponent by making his king checkmate. The game would however not be very interesting a pastime if you succeeded to do this as soon as you tried, but it would give you a very high level of usability. The intention the users, extended, action is to checkmate the opponent and the more effective this is done the better. The playability of chess is very high. It is close to impossible to fail in an intended interaction with the game. If the player wants to manipulate the table to move the bishop from G1 to E3, it is highly unlikely that the player's action will result in any other outcome. In the short run the two –ability's look very much alike, but it is when one sees to the context in which they are utilized that they become different. It never any risk that the two will collide, as they should never be used in the same medium with the one exception. A game should have high usability in the setup stage, before the actual game has begun. As this stage is not a part of the game structure it is all well to measure it in usability rather than playability. The player dose not wants to be challenged while trying to set up the game that would rather be very annoying unless it can be made to be a part of the playing experience itself.

III. Related Work: Models in Testing of Games

The GAP (Game approachability Principle) model is a heuristic list designed to highlight the awareness of playability- and approachability increasing factors in game design. Its main focus is to promote accessibility by good initial game state design and better tutorial design. To do this several issues have been identified that are of great importance for players when they first use a game. The GAP is based on several other models in the same field and with similar goals. Two of the models laying to ground for GAP are the HEP (Heuristic for Evaluating Playability) and the PLAY (Heuristic of Playability), which is a refinement of the HEP (Heuristics to evaluate playability).

Next to these two there are several other contributors that will not be mentioned further in this work but can be found in the HEP [15] and PLAY [17] articles.

The GAP (Game approachability Principle) model is a heuristic list designed to highlight the awareness of playability- and approachability increasing factors in game design. Its main focus is to promote accessibility by good initial game state design and better tutorial design. To do this several issues have been identified that are of great importance for players when they first use a game. The GAP is based on several other models in the same field and with similar goals. Two of the models laying to ground for GAP are the HEP (Heuristic for Evaluating Playability) and the PLAY (Heuristic of Playability), which is a refinement of the HEP (Heuristics to evaluate playability). Next to these two there are several other contributors that will not be mentioned further in this work but can be found in the HEP [16] and PLAY [17 articles. Below, the structural build of the three models are further described. The reason for this is that we will propose a further developed model, i.e. the GAID model. The structural build of that will be compared to those three models later in the paper.

A. The structural build of the HEP, GAP and PLAY models

The GAP model [8], [9], [10] is built on the foundation put down chiefly by HEP [15], PLAY [17] and the thoughts on learning through video- and computer games by Gee [18].

All these models address the four categorize that makes up a game design; game play, game story, game mechanics and game usability. In short these have the following meaning in the game context. The game play is the set of challenges that the player has to be overcome to beat the game. Game story is everything that is used to further the impression of reason for the game play challenge. Game mechanic is the rules by which the game play is solved/furthered. These are the limitations that define the action space. In good game design these should be in sync with the game story in such a way that the player understands the rules of the game without having to be explicitly told what the rules are. For instance, the it might not be clear to a player that never have been in contact with the Super Mario series before that by picking up (running over) a flower you all of a sudden get the ability to shoot fire. Very little about flowers can be connected through the everyday association tracks to being able to shoot fire. Hade it instead been a flamethrower it would have been more understandable that the rules of the game would change in the way of giving the ability to shoot fire. Game usability is about the interface of the game and any other means by which the player possibly can interact with the game.

HEP is a list of 43 item divided into four categories. The items are heuristics that are to give an indication of what to do but mostly what to avoid increasing playability in a game. The four categories are, just as in the other models, Game play, Game story, Mechanics and Usability. Among these 53 items 11 were transferred to the GAP list, the majority (six of them) coming from the Game play category. In the remaining 5 where 1 from Game story, 1 from Mechanics and the remaining were from usability. The six issues that were found in the game play subcategory

The PLAY is a refined list of the HEP, and therefore dose not offer much in change to add to the GAID. The Play uses the same heuristic items as the HEP, but adds 7 new to a total of 50 items. None of these new items are transferred to the GAP, and hence not to the GAID. In the PLAY the heuristics were further sorted than in the HEP. The PLAY had a total of three categories, being *Game play*, *Coolness/entertainment/humour/emotional immersion* and *Usability & Game mechanics*. These categories in turn had a total of 50 sub-categories. This shows that the former category of *Game story* from the HEP paper have been changed into the new *Coolness/entertainment/humour/emotional immersion* category. This wide category might be made in preparation for more items later on as all of the sub-categories in it at the moment only have one item in them. As the layout of the PLAY-list is done right now there is no benefit of having all the single item categories as a category by definition is grouping of phenomena.

The PLAY report is however strong in that it explains some parts of good game design that doesn't seam to have been able to fit into the list. One such, important part is to be vigilant of the difference between low usability and challenging game play. This should be included in some way in the GAID as it might be a tempting pitfall for game designers to fake challenging game play. Similar warnings have been made by game designer Ernst Adams in his lecture on bad game design named Bad Game Designer, No Twinkie! [7]. In this lecture Adams talks about some obvious examples of usability limitations camouflaged as challenging game play. Among these he mentions the inability to save games whenever the player wants to. This could be compared to not being able to save a word document whenever you want to. It is not something that actually enriches the game experience but rather something that could be sorted in under the number 2 bullet point "Players should not experience being penalized repeatedly or for the same failure". This leads up to one more bullet point for the GAID flowchart, which replaces the obsolete number 1 bullet "The first 10-20 minutes of play was fun and successful". The new heuristic should be "Did the player experience any difficulties with the interface or other parts that where not directly connected to the game play". If this question gives an affirmative answer it will lead to a category of problem that indicts that the design is experiencing a usability problem. The author justifies this by the description of usability compared versus playability that is given above. The reason to remove the first bullet point is because it can't be put in the hierarchy in any practical way. It will however be used as a stand-alone question to be posted when all the other questions in the list have been answered. This is a natural way to handle the number one question, as it in any other case becomes a kind of appendix to any of the other questions.

As it is designed now it will be a separate question from the flowchart that is put when any of the other questions fall into their categories as this indicates an error or bad design that needs tending to.

IV. THE GAID MODEL

Game Approachability Issue Definition (GAID) is an attempt to design a flowchart model based on the HEP, PLAY and GAP models. The goal of the flowchart is to be able to identify to which one of six commonly encountered problems in game design that a given problem is most likely to be connected to. The design takes a heavy stance in favour of the GAP model, as it is a combination of the other two models of HEP and PLAY. GAID is meant to be used as a tool to ease usability studies by taking the errors found in such studies and then identify the problem typ. GAID will however not, at present, be able to give the designer any guidance in *how* to solve the problem. This could possibly lie in a future development of the model into more detailed identification of the problems and also how these, more specifically identified problems can be handled. By reversing the GAID flowchart a R&D team backtrack through the chart to check if their product fulfils the points at all the junctions and by doing that possibly identify limitations in their design before usability testing have begun. If one branch has more errors than the other ones it can be an indication of where to focus more effort in the development process.

B. Reasoning behind GAID and its design

GAID is mainly an expansion on the GAP model. During the design of GAID it was judged that game approachability was the best aspect of design to focus on, as it is the focus of the GAP. Much also because of its importance to the success of game design, but also because of the combining of the other models that hade already been done by Desurvire et al.

In the GAP article [9] the authors look at the correlation between testing the GAP list as a heuristic and as a usability tool and also how this compared to the HEP and the PLAY in the same situation. From this it was concluded that the in both the case of heuristic testing and usability testing there were six recurring bullet points that was violated in all the tests. These bullet points where;

1. *The first 10-20 minutes of play was fun and successful.*
2. *Players should not experience being penalized repeatedly or for the same failure.*
3. *Varying activities and pacing during the game in order to minimize fatigue or boredom.*
4. *The game provides clear goals; overriding goals are presented early and short-term goals throughout game play.*
5. *The skills needed to attain goals were taught early enough to play or use later, or right before the new skill was needed.*
6. *The game gave rewards that immersed the player more deeply in the game by increasing their capability or capacity to do things in the game.*

Based on these six bullet points the GAID flowchart has been made. When the six bullet points were checked against their categorization in the HEP- and the PLAY lists it became clear that they where all categorized as *Game play* issues. The main difference between the two models was that the PLAY list hade made several subcategories to *Game play*. Three were categorized as *Goal* oriented heuristics, two as *enduring play* and the last as *variety of players and game style*. This already clear division of categories was used to arrange the hierarchical levels in the flowchart. By doing this, the most important item could be at the highest level in the flowchart and then the rest in a descending fashion. If the first issue in the flowchart was not fulfilled it would not mater if the ones of lower importance would be fulfilled or not as the more important issue would have higher priority to be solved any way. The problems categorized to be of *Goal* nature was decided to be the most important problems as without knowledge of the goal of action, action becomes void. Second in importance was deemed to be the *Enduring play* items. For approachability to be made successful the participant must give the game a chance (although it falls on the game to instil the trust needed for this to happen) to show its qualities. Last was the *variety of players and game style.*

Through the study of the report on PLAY the categorization of the first of the six issues as a game play issue became questioned by the author of GAID. It later became removed from the GAID flowchart because of its inability to be integrated in the flow in a practical way. A new issue that differs between playability issues and usability issues instead replaced it. Something that otherwise can be a dangerous pitfall in game design.

C. The GAID flowchart

The GAID flowchart is designed with the most urgent questions at the top and issues then following in a descending fashion though the chart. Every diamond corresponds to one of the six issues from the above section with the exception of the first issue that was instead added as an appendix in the end of the chart. First, the flowchart is presented and then further described below.

In the flowchart above, diamond 1 corresponds to the added question about usability versus playability. This is put as the topmost question as it is flows into the rest of the questions in the case of a negative response. Otherwise it shows that the problem is of a usability nature.

This is not the focus of this model and therefore it is important to get it out of the design process early on. For this outcome it is recommended to delegate the problem to a software developer and get the problem away from the game designer's sphere, it is not her responsibility.

Further, the second diamond is the most important question in the game design sphere. It responds to the forth issue in the above section, i.e. the players' knowledge of the goal.

To know what the purpose of a game is the first thing that the player must learn. First when this is clear is it worth to learn anything else. It might be argued that the enjoyment of the game is the goal of any game session in itself. This argument gets acknowledgment from the author, but a game of any higher sophistication will most likely (speculation from the author) will lose its entertainment value if it is not played in the intended way or at least perform at a non-optimal level of entertainment. This can be exemplified by the simple dice game Yatzy. The goal of the game is to gather points unless this is known it will be hard to get any entertainment out of the game as it will be at most some random rolling of dice till someone else finally (most likely) wins.

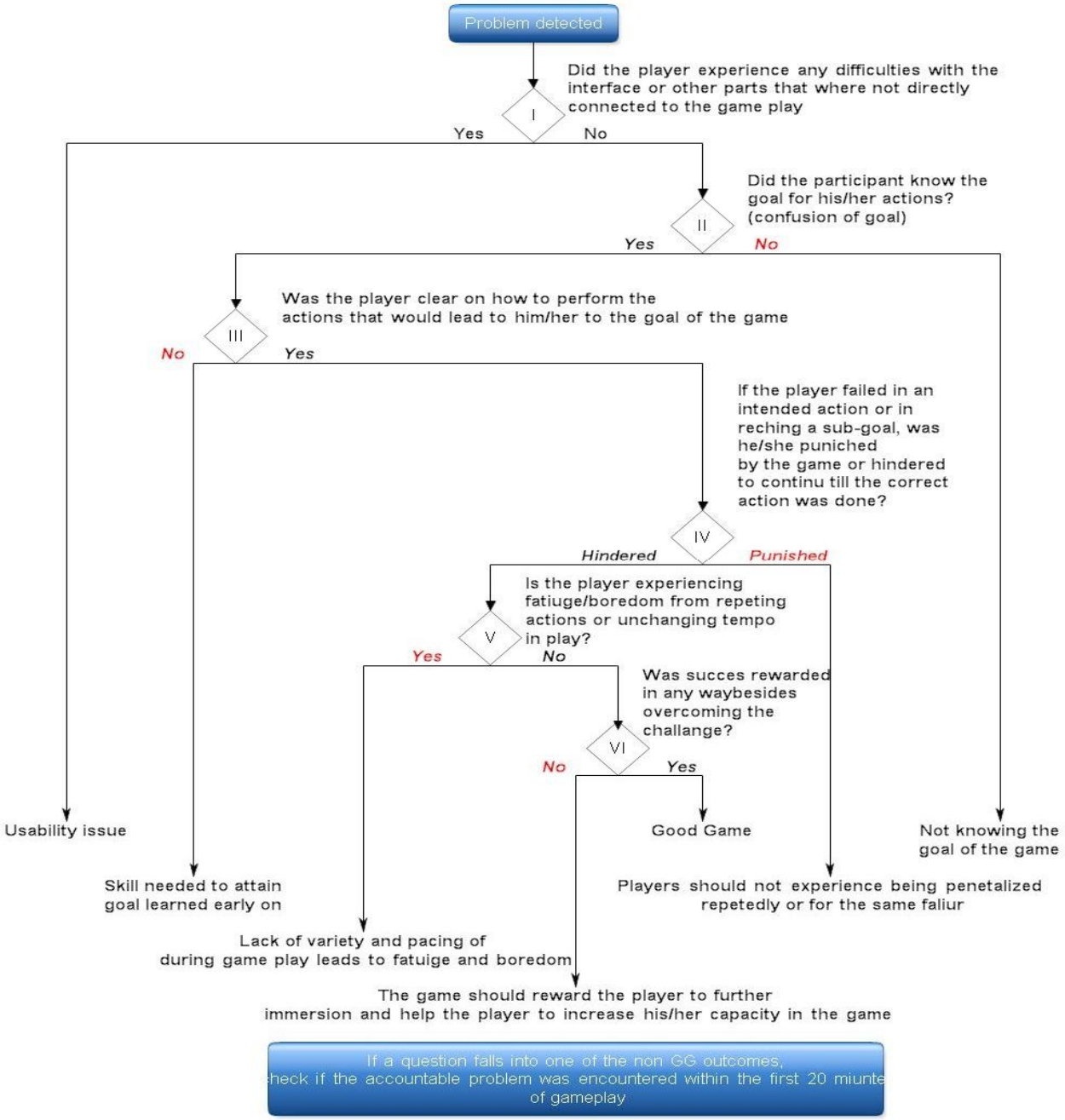

Fig. 1 The GAID Flowchart

The third diamond corresponds to the fifth issue in the above section, i.e. the need to learn skills before they are to be exercised in a real game situation. By the example with the Yatzy game this would be the skill to throw the dice or perhaps even the understanding of what score to go for at a given result of the first role of the dice. This is put as the second most important issue for a successful game design and possibility to approachability as any game were the player do not understand how to reach a desired goal or sub-goal very soon will become frustrating or boring if not both. This problem nature corresponds strongly towards playability and possibly also towards design of interface, which in the case of non-entertainment software often is the reason for a users inability to perform an intended action.

The forth diamond is quite specific in its query. The question corresponds to the second issue in the list above, "Players should not experience being penalized repeatedly or for the same failure". This question is tricky to answer from just making an observation of a player's behaviour on a video recording. It is possible to make informed assumptions, even when the user of the GAID is not an expert, that if an annoyed appearance or language is displayed in connection to a series, and especially triggered by, being punished at a failure to overcome a game play challenge it would fall under the outcome of *punished*. To approach this problem the designer might want to consider *fencing* to hinder the player from making all to big errors. This might however have the undesired effect of robbing the player of the sense of achievement when the though challenge is finally overcome if it is obvious that it was more due to the helping hand of the game than the player own skill. Some players' play exclusively for the bragging rights of having overcome a specifically hard challenge without ever failing such as Blizzard Entertainments classic Diablo 2 played in hard-core mode [19].

The fifth diamond corresponds to the third heuristic in the above section, "Varying activities and pacing during the game in order to minimize fatigue or boredom". This issue is about the staying power of a game and responds to the players desire to experience new things. If a game can be made up in such a way that the player never experiences that he/she have to repeat some tedious action (or do one first hand) fatigue and boredom can be avoided. It is however understood that to build tension or any other feeling of suspension to finally relive it in a climax, less intent or even close to boring sections might have to be implemented. To avoid complete boredom it is recommended to mix different spheres of interest so that when one specific kind of tension is relived the player is already building on some other kind of tension without even knowing it. This model to avoid boredom has been greatly exploited in the massive multiplayer online Role playing game genre (MMO). In the MMO *Warhammer online* the player collects *experience-points* to level up the character/avatar he/she is playing but at the same time there is a collection of *renowned-points*. The system is set up in such a fashion that when a player makes a character level the next renowned level is never fare from being cleared and when that is done it is close to the next character level. By this system there is almost always something that can spur (Gee, 2008) on the player to keep playing. The reward is never far off.

This issue is placed as the fourth game plays issue in the hierarchy due to the fact that even though important, staying power can be achieved even though a game has a very repetitive mode of game play. The author will once again refer to the Blizzard entertainment classic Diablo 2 that have been keeping its players since April 2000 when it was released till today. The game is extremely simple and only offers any real variation in what means the player choose to use to kill his/her way through never-ending waves of generic monsters. This game has in some way been able to spellbind its audience despite being extremely repetitive in its game play. By this it seems like even games that relies on a very limited repertoire can have high staying power.

The last diamond corresponds to the sixth issue in the above list. This issue answers to the need of motivation the player. The heuristic asks for items that might motivate the player to overcome sub-goals and at the same time increase the immersion, capacity and/or capability to perform in the game. Items fitting into this category are traditionally power-ups and new skills but can also be new pieces of information in games with strong story drive. This issue goes hand in hand with diamond 3 that asks if the player have the proper knowledge to fully benefit from the possibilities that the game presents him/her with.

If the flowchart runs its full course and passes all of the six diamonds without taking any of the red paths the problem is not of any of the six most common types that have been identified on the GAID model. In this case it is recommended that the specific case be delegated to a more experienced usability expert or game designer who can move outside the framework of a model.

After any run of the flowchart in which a result emanates in one of the six possible problem categories or into the catch condition of delegating the problem to a more experienced usability expert or game designer, the user of the GAID must put the appendix question. This question corresponds to the first issue in the above section, "The first 10-20 minutes of play was fun and successful". This is of great importance as this seems to be the critical period in which a new player decides if the game is worth investing time in or not [17]. The reason for putting this in the end of the flowchart is that the condition of a problem to appear during a specific time interval doesn't really categorize it as any specific type of problem. It will however add to the severity of the problem if appears during the critical time period. If the problem appears within the first 20 minutes of play, the problem should be prioritized over any problem of equal severity that appeared after the critical period.

In the end the flowchart has seven possible outcomes, of which five are categorized problems within the realm of game design, one outcome is of general usability type and the last one a catch up function for anything that can not be fitted in the above.

V. PROCEDURE

In this study a flowchart-model was made to support design of computer games by simplifying usability studies and initial design. The model takes a special interest in approachability, as the name implies (Game Approachability Issue Definition). This is inherited from the GAP model which is the main parent model together with PLAY and HEP. From a cross-model test involving all of the parent models six issues were identified as being recurring in all cases of the tested usability studies and heuristics.

From these six issues five were used as identifier questions in the flowchart and one was used as an appendix question when it in itself did not categorize a *type* of problem but rather the severity of the problem at hand. One more question was added to differ between usability issues and playability issues.

To verify that the issues that were picked for the GAID flowchart had any validity in the to the interest of the users a several usability tests were observed from an open ended study that was made before the GAID model was designed. An introspective study was also made to check for presence of the six issues in a tutorial environment.

A. The usability study

In the usability study a sample of 137 participants aging from 6 to 16 (average about 10,5). Of which 49 were girls and 88 boys. This distribution between the sexes is representative for the gaming community in general, which has a 60♂ - 40♀ distribution between men and women while this sample has a 64♂ - 36♀ distribution [20]. The sample was drawn from the junior and youth teams of a local soccer club. These participants were to represent average young users of console/computer games. The participants conducted the study per team. This gave us a natural age and gender grouping. The smallest groups to do the study were five participants and the largest where 19 participants. Over the entire study a grand total of 206 observations were made as 69 of the participants returned for the voluntary follow up study.

The data gathering of the study was divided into two phases, much like each other but with different games and an additional hardware in the second phase. The reason for the division was to see possible trends and to gather a larger quantity of data from the sample. After having conducted the first session the participants were asked if they would be interested to participate in an additional session at a later date. The ones that showed interest were called for the next phase of the data gathering. The procedure for the second phase was the same as for the first one with the exception that other games were used. Some testing with an IPhone was also done.

The test environment was set up in the clubhouse of the soccer club that the test subjects where members of. This was mainly done as a convenience matter to the. Further, the close proximity of the clubhouse to many of the participants' homes made it more viable for the participants to get to the experiment on their own. Something that might be a limiting factor when working with young participants. We also saw the use of a familiar location to be favourable as it might increase the ecological validity of the data gathering. If the data gathering had been made in a laboratory environment it is highly possible that the participants might have been distracted or uncomfortable. This might be especially important to consider when using young participants that otherwise might become worried by unfamiliar surroundings such as a lab can be.

The setup of the lab in phase one of the data gathering was that of three stations. Two stations were equipped with consoles, one Microsoft Xbox 360 and one with a Nintendo Wii. The third station was a computer. The two console stations were equipped with microphones and dual cameras aimed to capture both the participants playing the game and the monitor showing the game. The camera aimed at the participants where positioned so to get a good view of the controller being manipulated by the user. The third station lacked surveillance in the form of recording as it was initially intended to be an offloading station for groups of more than 12 participants. It would later in the data gathering however show that the information that was observed by test leaders at this station was quite interesting. In phase two an IPhone was added to the station cycle. This was made as an exploratory observation to see how the participants used gestural based interfaces such as are frequent in IPhone games. Whit this additional station the number needed for the computer offloading station was increased to 17 participants. None of this data was used in this study.

The games that were tested on in the first session were the following. On the XBOX 360; Virtua Tennis 2009, Banjo Kazooie: Nuts and Bolts, Kung Fu Panda, FUEL, Spiderman: friend or foe. On the Nintendo WII; Indian Jones: The original adventure, Punch Out!!, Pokemon Battle Revolution, Super Paper Mario WII, WII sports, Mario kart WII

In the second session some new games were added to the supply. The new XBOX 360 games were; Beautiful Katamari, Ice Age Dawn of the Dinosaurs, Fifa 09 360, Forza2 motorsport 360. The WII games in the second session were; Wii sport resort, Wario shake dimension wii, Lego star wars the complete saga and WII sports.

The recording was made using two Fujitsu Siemens computer laptops with ADM Athlon(tm) X2 Dual-core QL-62 2.00 GHz running Windows Vista home edition, running Morae Recorder version 3.0.0 as recording software. One of the computers where connected to a Blue microphones model Snowflake and the other to a Samson C01U USB microphone. One computer used a Logitech Quickcam E 3500 to record the participants playing. The other computer used a Logitech Quickcam Pro 9000 to record the participants.

Both recording computers used their built in Amilocams' to record the screen where the game was being played. The actions of the participants where also observed by two test leaders who took notes on interesting behaviour and quotes. These notes were to be used as a support to the recorded material when looking for interesting behaviour in the study.

The first thing that the participants got to do was a questioner[1] about their gaming habits and personal information needed for filing. The questioner also asked about what consoles they had in their home as it was suspected that such a presence could influence the performance in gaming and understanding of games played on the same console in the study.

The groups where evenly distributed over the stations, with the two console stations being filled up first in the case of less than sufficient participants to fill all three.

The participants then got to play at their designated console for 30 minutes. During this time they got to play in any fashion they wanted whit the one rule that everyone hade to play. They where also encouraged to play games of which they had no experience. They where also urged on to solve problems without asking the test leaders for help but rather to solve them as if there where no one grown up to ask. It was allowed to talk within the own group about how best to solve a problem. If a problem experienced by the participants, was judged to be outside the scope of what the group could solve by them the test leaders would intervene to let the process continue.

After about 30 minutes the groups swapped stations. In between the changes the participants' were served lemonade and fruit. This was done to further the stamina of the participants who in other case might have lost interest in the exercise. After the break the groups changed stations with each other. In the case with three stations two changes where made after the break.

The data gathering ended with a debriefing about the payment that the teams where to get in exchange for their help and during phase 1 an inquiry about the participants further interest to take part in phase 2.

This study hopes to counter for bias by not having any actual intention when it was made, but rather tries to be as qualitative as possible so to be able to be usable in many different studies later on. By recording as many factors as possible it will be able to be used in many different studies.

B. Limitations of the usability study

The age of the participants is not representative for the general user group of entertainment software, where the average user is 35 years old and has played games for 12 years.

Second is the concern that some of the participants might have been exhorted to participate in the study by their parents or coaches. To counter for this bias the teams were asked to only send children who wanted to go by their own will. Even though this precaution was made, at least three obvious cases of children who where participating against their own interest was made by test leaders. In these cases the child in question was excused from playing, as this would not be defendable form a research ethical point of view [21].

The last limitation that has been identified in the usability study is the possibility that there might have been to few test leaders to make a qualitative data gathering. In the work of Tulahimutte and Bolt it was suggested that in game usability studies of similar model to the one conducted by us a ratio of 1:6 test leaders to participants was acceptable [22]. Even though all the session that were conducted were recorded so that no material should be lost the material becomes much more accessible when using the supporting notes of the test leaders.

C. The introspective study

The use methods where the researcher makes an active self-reflection of an experience can be quick and powerful methods [23]. In the introspective study the author of this paper played through the tutorial level of Heroes of Newerth. During the tutorial the appearance of the issues that make up the GAID were noted down. It was found that all the issues from the GAID could be fond in the tutorial. This gave good indications that the issues addressed in the GAID were representative for informed approachability in a well-designed tutorial. The test was run on a Intel(R) Core(TM)2 CPU 6300 @ 1,86GHz 1,87GHz 2GB RAM. The version of Heroes of newerth was beta 2.64.

D. Limitations of the introspective study

As with all introspective studies it is impossible to copy the exact procedure that took place. This is not however to much of a problem as the study was made to give an indication rather than a precise answer to the question; are the issues in the GAID representative for good approachability.

E. Validation of GAID

To validate the GAID model two tests were made. One introspective investigation made towards the tutorial level of the online real time strategy game *Heroes of Newerth* [24] and a post-test application of the GAID on a video- and observation material made on 139 participants playing video games under monitored forms.

In the introspective analysis was made in the tutorial environment of Heroes of newerth where the six issues introduced above were sought after. The entire session was done in about 20 minutes, which is within the margin of the first bullet point, which states that the critical period that the game has to establish interests and liking is 10-20 minutes of enjoyment and success. To introduce the game to the player for a longer period than necessary makes the risk of the situation becoming boring. The first thing that was addressed in the game was the *Goal*.

This gives support to the hierarchical placing of *Goal* in the GAID flowchart. Without knowledge of the goal all other knowledge is void. It is therefore the most prioritised information to give to the player. Next was introduced how to obtain the goal by several sub goals. This was identified to be the equivalent of the skills needed to attain goal bullet point. After this the instructions of what actions that was needed to attain the sub goals, and in the long run the overall goal. This was also identified as a form of skill needed to master the game. After the instructions of basic game play were given rewards were introduced in the form of items and additional skills. This could be identified both as a form of reward like bullet number 6 but also as variety in pacing and activity. During the entire tutorial there was very heavy fencing to hinder the player from failing in their learning. This is not really a part of the 6 issues but it helps to avoid penalization of players during the all-important first 10 - 20 minutes. In the beginning of the tutorial for example it was impossible to die with the character if you did not kill all the little enemies that were spawned to help in the instruction of how to kill the opposing team. It was also made impossible to buy the wrong items during the buying instruction sequence.

In conclusion to the introspective study it was found that all the 6 of the bullet points that were found to be of importance to game development and approachability could be found represented in the tutorial session of Heroes of Newerth.

The second validation of the GAID was done by looking at an extensive user test with 139 participants that was done without any specific intention in mind besides getting a large, qualitative observational data for usability research. By observing the participants playing in an uncontrolled, but yet observed environment the parts that makes up the GAID flowchart can be supported by quotations from the user sessions and by research leaders observations.

The following observations and quotes were found and picked out in support of the issues that make up the GAID.

The *Usability* versus playability issue, the first issue in GAID, is supported as being an issue that is of interest by the following example that was observed when two boys from the participant group of boys age 8 were trying to set up the game pokemon on the Wii console.

Participant 1 (P1): *check it out, it doesn't work!*
Test leader (TL): *what are you trying to do?*
P1: *check it, when I press that one, that pokemon. I still don't get him!*

The participant hade obvious problems with the setup due to lacking usability design. In this case the test leader observed that a problem was happening and made an inquiry to investigate, but the issue could still have been identified by the GAID as a usability problem as it happened before actual game play hade begun.

To support the *Goal* issue the following observation was made in the usability study. The fist was made with girls age 15 playing Fifa 09 on XBOX when they were observed trying to identify a sub-goal in the game. After scoring a goal in the game it is possible to press a certain combination to perform a goal gesture. It has to be done when the instruction to do it appears, a so-called quick time event (QTE).

P1: *What should I do?*
P2: *Then it should be like that, a goal gesture you know*
P1: but why R S?

This could be identified as either a goal issue or a skill issue but as there seamed to be a confusion on what the action of pressing the "R" "S" combination would result in it was decided to be a *Goal* issue first

Two observations were made to support the *Skill* issue. The first was made with participants from boys' age 13 shooting hoops in Nintendo Wii sport resort.

P1: *you should just shoot you know. Do you know how to do it?*
P2: No..?
P1: *You go down and press B then you throw it…*
P1: man this is not anything like shooting hoops. It doesn't work!

By the GAID this was identified as a *skill* issue as P2 clearly states that he doesn't know how to solve the sub-goal of shooting the ball in the hoop.

The other observation was made with girls age 16 playing Fifa 09 on XBOX 360. Once again it is a clear statement of lacking knowledge from one participant to another.

P1: *Play game*
P2: *no! I never know how these things work!*
P1: *It's all about pressing buttons.*

It seemed like neither p1 or p2 hade any real knowledge of what skills were needed to solve the goal of the game.

The issue of Punishment was hard to support even from the huge material available from the usability study. The one instance that was found that could be identified, as a possible punishment for failure was when one participant of boy age 10 was knocked out in Wii Punch Out!! and had expressed clear dismay when he had to watch the looser cut scene before getting to have another try.

P1: *Ah man! Now we have to wait like a year again!*

The issue of *Tempo* is represented by a quotation from girls age 15 playing Lego Star wars on Wii after they have been trying to further the game play in 5 minutes without getting anywhere. The participants have reached a dead end and cant get anywhere.

P1: *All right. Now we can change game cause I don't get it anymore. I don't know what we are doing and that is worthless.*

During the same star wars session there was a fine example of the *Reward* issue being observed when the two participants played Lego Star wars on Wii.

P1: *Wow! Check it out! You get life back when you kill stuff*

P2: *Nice!*

This is a clear case of a reward for solving a sub-goal. Also, even if it was not said out loud in the quotation above the life that was gained from killing opponents was also mixed with money that gives the player the possibility to by power-ups, something that was made later on by the same participants.

The appendix question concerning critical time was observed several times in the usability material. One very clear quotation was found when participants of girls' age 8-9 were playing Mario kart Wii. (the name of the participant is not the real name of the participant that was in the study)

P1: *Its not working out so good for you, is it Mary? Turn! Turn! Turn!*

P2: *I don't want to play this game!*

Mary: *what gives!? You can't drive this stuff!*

In conclusion from the usability study it is clear that all of the issues that are represented in the GAID are often issues that appear in game play situations. Some of the issues are suggestions to ease the approachability of a game while others are checkpoints for issues that have to work properly to get a good game play experience.

VI. CONCLUSIONS

In this work we have tried to improve on the GAP model designed by Desurvire & Wiberg [9] [10]. The product of this work is the Game approachability issue Definition, GAID, a flowchart model that uses the most commonly occurring issues in the combined approachability models, GAP, HEP and PLAY. To these issues was also added the issue of playability being confused with usability. One of the original six issues from the GAP list was removed from the flowchart and added as an appendix question that could adjust the severity of the problem being identified.

To verify the accuracy of the question in the GAID model an introspective study was made while playing the tutorial level of a real time strategy game. Next to this an observational usability study of 139 participants was made to support the issues in the GAID. As the material from the usability study is about 42 hours worth of recordings it is here admitted that all of the material has not been taken part of in this study. Rather, some parts that have been selected from the observations and notes that were made by the test leader during the data gathering and by this selected parts of the recordings have been given special attention in the search for supporting material for the GAID issues.

To be able to judge the outcome of the GAID it would have to be field tested in a real life design situation, but from the testing that have been made by the author a promising prospect is seen. GAID takes the most important issues from some very powerful model and use them to identify common shortcomings and to highlight important parts needed to create working accessibility. With the usability study that was used for the evaluation of the issues, a more specific approachability focused on children could be of interest to the computer game community. This might be done if further work is done on the GAID flowchart.

ACKNOWLEDGMENT

We would like to thank all of the participants in the study for their efforts and patience. It was so inspiring to find out how you approach games in your life. Further, we would like to thank our colleagues for valuable input in earlier versions of this paper.

REFERENCES

[2] C. Wiberg, K. Jegers, and J. Bodén. "Cross media interaction design." CHI'07; Workshop HCI and New Media Arts: Technology and Evaluation, April 28 – May 3, San Jose, CA USA 2007

[3] H. Desurvire, and C. Wiberg. "Evaluating user experience and other lies in evaluating games." *Computer Human Interaction Conference*. 2008.

[4] H. Desurvire, K. Jegers, and C. Wiberg. "Evaluating fun and entertainment: Developing a conceptual framework design of evaluation methods." *Facing Emotions: Responsible experiential design INTERACT 2007 conference, Rio, Brasil*. 2007.

[5] C. Wiberg. "From ease of use to fun of use: Usability evaluation guidelines for testing entertainment web sites." *In Proceedings of Conference on Affective Human Factors Design*. 2001.

[6] C. Wiberg. "Usability and fun: An overview of relevant research in the HCI community." *Proceedings of the CHI Workshop on Innovative Approaches to Evaluating Affective Interfaces*. 2005.

[7] C.Wiberg. *A measure of fun: Extending the scope of web usability.* Department of informatics, Umeå University. 2003.
[8] E.Adams,"www.designersnotebook.com."*Lectures.*1252010.http://www.designersnotebook.com/Lectures/lectures.htm (Used 18 5 2010).
[9] H. Desurvire, and C. Wiberg. "Master of the game: assessing approachability in future game design." *CHI '08: CHI '08 extended abstracts on Human factors in computing systems.* Florance, Italy: ACM, 2008. 3177-3182.
[10] H.Desurvire, and C. Wiberg. "User Experience Design for Inexperienced Gamers: GAP – Game Approachability Principles." in *Evaluating User Experience in Games*, Regina Bernhaupt (ed.), pp. 135-152. London: Springer, 2010.
[11] H. Desurvire, and C. Wiberg. "User Experience Design for Inexperienced Gamers: GAP–Game Approachability Principles. *Game User Experience Evaluation*, Bernhaupt, R. (ed.) 2015.
[12] W. Barendregt, B. Mathilde and S. Mathilde. "Empirical evaluation of usability and fun in computer games for children." *INTERACT 03*, 2003: 705-708.
[13] T.W. Malone. "Heuristics for designing enjoyable user interfaces: Lessons from computer games." in *CHI '82: Proceedings of the 1982 conference on Human factors in computing systems*, by ACM, 63--68. New York, NY: ACM, 1982.
[14] M. A. Federoff. *Heuristics and usability guidelines for the creation and evaluation of fun in video games.* PhD report, Cite seer, 2002
[15] Preece, Y. Rogers, H. Sharp, D. Benyon, S. Holland & T. Carey. *Interaction design: beyond human-computer interaction.* Wiley; 2 edition, 2007
[16] H. Desurvire, M.Caplan, and J.A. Tooth. "Using heuristics to evaluate the playability of games." *CHI '04: CHI '04 extended abstracts on Human factors in computing systems.* New York: ACM, 2004. 1509-1512.
[17] C. Fabricatore, M. Nussbaum and R. Rosas. *Playability in Action Videogames: A Qualitative Design Model.* Vol. 17, in *Human-Computer Interaction*, 311 — 368. Naples: Taylor & Francis, 2009.
[18] H. Desurvire, and C. Wiberg. "Game Usability Heuristics (PLAY) for Evaluating and Designing Better Games: The Next Iteration." *Online Communities and Social Computing.* Berlin: Springer-Verlag, 2009. 557-566.
[19] J. P. Gee. *What Video Games Have to Teach Us About Learning and Literacy.* Palgrave Macmillan, 2008.
[20] Blizzard Entertainment. *Blizzard entertainment Diablo 2.* 27 05 2010. http://us.blizzard.com/en-us/games/d2/ (Used27 05 2010).
[21] The Entertainment Software Association. "www.theesa.com." *The Entertainment Software Association - Game Player Data.* 5 May 2010. http://www.theesa.com/facts/gameplayer.asp (Used5 May 2010).
[22] Vetenskapsrådet. *Forskning som involverar barn.* den 20 5 2010. http://www.codex.uu.se/manniska1.shtml (Used20 5 2010).
[23] T. Tulathimutte and N. Bolt. "The science of fun: One-to-many moderated game research." *CHI 2008.* Florance: ACM, 2008. 2033-2038.
[24] P. Wright and J. McCarthy. "Empathy and experience in HCI." *Proc. of CHI '08.* New York, NY: ACM, 2008. 637-646.
[25] S2 Games. *Heroes of Newearth.* 14 15 2010. www.heroesofnewerth.com (Used14 2010 2010).