

Analysis and Comparison of Software-Tools for Cognitive Assessment

Leslie Tso

Universität Stuttgart
Universitätsstraße 38
D-70569 Stuttgart

inf82940@stud.uni-stuttgart.de

Christos Papagrigoriou

Universität Stuttgart
Universitätsstraße 38
D-70569 Stuttgart

swt76716@stud.uni-stuttgart.de

Yannic Sowoidnich

Universität Stuttgart
Universitätsstraße 38
D-70569 Stuttgart

swt87611@stud.uni-stuttgart.de

ABSTRACT – GERMAN

Aufgrund einer stetig steigenden Anzahl von eingeschränkten (und älteren) Menschen ist es notwendig geworden, eine Methode zu finden, diese effizient in die Arbeitswelt (bzw. Gesellschaft) zu integrieren. Diese Arbeit legt seinen Fokus auf die Analyse und den Vergleich von drei Software-tools, welche die kognitive Fähigkeit von leistungseingeschränkten Personen messen soll. Die drei Werkzeuge nennen sich GATRAS, entwickelt von der Universität Stuttgart, CogState von CogState Research und die computerbasierten Tests von hamet e entwickelt vom Berufsbildungswerk Waiblingen. Die Arbeit soll eine detaillierte Beschreibung der drei genannten Tools geben, sowie diese einem genauen Vergleich unterziehen. Zusätzlich wurden die Tools alle in einer Studie mit 20 Teilnehmern getestet, in Zusammenarbeit mit der Gemeinnützige Werkstätten und Wohnstätten GmbH (kurz: GWW) in Sindelfingen. Durch zeitliche Beschränkungen konnten jedoch nicht alle „Spiele“ der unterschiedlichen Software-tools getestet werden, deswegen wurden nur die von den jeweiligen Firmen empfohlenen Spiele überprüft.

ABSTRACT - ENGLISH

Due to the rising number of impaired and elder persons, it has become crucial that we find methods where we can easily and quickly integrate them into the workforce and by extension, society. This paper focuses on the analysis and comparison of three software-tools that assess the cognitive ability of people with impairments. The three software-tools are GATRAS by the University of Stuttgart, CogState by CogState Research and the computer-based tests from the hamet e by the Berufsbildungswerk Waiblingen. This paper will give a detailed description and comparison of each software and their features. In addition, the software-tools will be tested in a study with 20 participants in conjunction with the Gemeinnützige Werkstätten und Wohnstätten GmbH (GWW) in Sindelfingen. However, due to time constraints, not all games will be tested but the recommended battery of tests from each software will be used for the study.

INTRODUCTION

What are software tools for cognitive assessment? Why do we need it? Software tools for cognitive assessment are computerized tests which evaluate the cognitive abilities of people with impairments. Today, there are many reasons why we need access to such software-tools. One of the

main reasons is due to the growing number of people with impairments. According to the World Health Survey done by the World Health Organization in 2011, over 15% of all people over the age of 15 suffer from a disability whereas the Global Burden of Diseases estimates the percentage to be as high as 20% [1]. Today, those numbers add up to between 1 and 1.46 billion people. Due to these large numbers and anti-discrimination laws in many countries such as the “Equality Act 2010” in the United Kingdom [2], it has become essential for companies and governments to induct people with disabilities into the current workforce. Furthermore, due to several factors such as the weak global economy [3], longer average life expectancy and low fertility rates in countries such as those in the European Union [4], many pensioners and elderly find themselves having to return back to workforce or delay their retirement. In fact, in 2008, there were more people over the age of 65 than under the age of 15 in the European Union [4]. Computerized cognitive assessment tools are a way to help integrate impaired and elderly people into the everyday workforce quickly and efficiently. As time is often not a luxury that many companies have, software-tools may be the solution to intensive and time-consuming interviews and tests for people with varying types and degrees of impairment.

Throughout the course of the study, the software-tools will be evaluated based on several factors. One factor by which the assessment tools could be evaluated on is the “usability” for both the users and the testers. Under the definition of “usability” by Nielsen [5], the tools will be evaluated on:

- How easy it is to use each individual software for the first time
- How quickly or efficiently the user can accomplish the given task
- The ease of remembering the controls and functions of the software
- How easy it is for the user to make mistakes
- Whether or not the users are happy with the software in general.

Furthermore, as the testers focus more on the results received from the software rather than the process of the software itself, the functionality and capability of the assessment tools such as automatic score analyzers and the readability of the results will be analyzed and compared.

In the first step the tools' tests will be analyzed. However, as this study focuses on assessing the cognitive abilities of people with varying cognitive disabilities, test batteries (a set of tests with a common scoring method which evaluates a specific treatment group) will be chosen from each software that fits the requirements. For this study, all tests that assess cognitive function in GATRAS will be used as there are only five tests that focus on assessing the cognitive function of participants. For CogState, the "Early Phase Battery" recommended by CogState will be used as it tests a wide range of cognitive functions. Furthermore, the Level 2 computerized tests from hamet e will also be analyzed as they were generally created with entry level tests in mind.

In a second step we will analyze the advantages and disadvantages of each software-tool. In this step, factors such as the software-tool's usability, features and user satisfaction will be taken into account.

In a third step we will analyze test data from a pre-study with participants from the University of Stuttgart and a main study with participants from the Gemeinnützige Werkstätten und Wohnstätten GmbH. In this step, feedback will be received from the participants on the usability of each of the three softwares as well as their general satisfaction of the product. Furthermore, the test data will be compared with each other in order to see if the tests from each software correlate with one another in order to establish a common baseline. Additionally, in the case of the main study, the results of each software-tool will be compared with the results from the Ground Truth Study, a real world test in order to see whether the results from the software tests can be useful in estimating the results from real world activities.

Based on the results from the evaluation of the individual tests, the advantages and disadvantages of each assessment software and results from the analysis of the test data from the studies, a recommendation for future use and required developments will be made.

BACKGROUND

Gamification

Each of the three software-tools analyzed in the paper invests heavily in the "gamification" concept. But what is "gamification" in the first place? According to the Oxford Dictionaries, "gamification" is "the application of typical elements of game playing (e.g. point scoring, competition with others, rules of play) to other areas of activity, typically as an online marketing technique to encourage engagement with a product or service" [7]. In other words, it is the use of game elements such as designs and techniques for non-gaming applications such as tests or filling out forms. "Gamification" did not gain widespread popularity until the rise of easily accessible video games and internet connectivity [7]. However, it has been in use for much longer. A good example of this is frequent flyer

programs from airlines [7] which have been in use since 1979 [8]. The frequent flyer programs allow members of the airline program to exchange "air miles" for certain rewards such as cheaper or free tickets or free seat upgrades much like the "points" and "achievement" system used in video games today. Furthermore, the frequent flyer programs allow their members to "level up" [7] depending on how many miles they accrue yearly where each level allows the member to have more benefits such as the use of the lounge or heavier luggage. This is similar to gaining experience in video games today in order to level up and be allowed to use better equipment or have better attributes. Another method in which frequent flyer programs use "gamification" is with their challenges such as "Fly 3 segments in the next 90 days for 2500 bonus miles." [7]. This is very similar to "quests" in standard role-playing games where the user has to complete a certain task such as "slay 10 wolves" in order to receive a certain reward such as gold and experience or certain items. However, it is not only airlines that have been using "gamification" in order to promote their businesses. An example would be many of the fast food restaurants and cafés. Many fast food restaurants give their customers a stamp for every meal they order and after the customer has accumulated a certain number of stamps, they are given a free meal or gift.

"Gamification" comes with several benefits for users as well as companies and organizations which implement it with their activities. With the majority of children and young adults having grown up in an environment surrounded by video games on their computers, televisions or even cellphones, "gamification" is a very effective method for engaging said people in mundane or repetitive tasks. Statistics from today's youth support this. Currently, 58% of all Americans play video games and the average gamer today is 30 years old and has been playing games for nearly half their life [9]. This means that people joining the workforce would already be well versed in the intricacies of gaming and would have little problem integrating into a "gamified" workforce. Furthermore, with the implementation of scoring systems and achievements, naturally competitive users may feel more motivated to do more and better. In addition, even naturally uncompetitive users will feel a sense of accomplishment if they achieve a specific achievement or get a very high score. "Gamification" can also be used as a tool to compare the user to his or her colleagues and to see in what areas he or she should improve in as well as easily see the rate in which they are improving. Furthermore, coupled with real world rewards, "gamification" could further increase a user's motivation and dedication to his or her job. Last of all, when the user is having fun, he or she would not object to working or "playing" a little longer.

Despite the many advantages of "gamification", it does come with several drawbacks. For example, a "gamified" environment could be difficult for people who did not grow up in environments constantly surrounded by video games

such as elderly people who grew up before the popularity and development of video games as well as people who grew up in environments where the majority of the community could not afford devices such as computers or smartphones. Furthermore, “gamification” may cause unnecessary stress for people who do not work well under pressure or dislike competition or comparisons to other people. In addition, if a user sees that he or she is doing badly in comparison to his or her colleagues or in general, he or she might become depressed and lose motivation. Additionally, as “gamification” is still a relatively new concept, a standard or ideal design is still not available. As there are a wide range of game elements that could be implemented into work processes, users that work well with a particular element are not guaranteed to work well with other game elements. Also, negative achievements could be disadvantageous to the work environment whether it is implemented or not. On one hand, if there are no negative achievements, users might try just to get as many achievements as possible without checking the quality of their work. An example would be managing to fix 100 bugs in a program in a month but with unorganized and uncommented pieces of code. On the other hand, if negative achievements are implemented, a user may feel unhappy or unsatisfied and lose their enthusiasm if they get one [10].

Commonly Used Game Elements

While today’s games consists of hundreds of elements which work together to engage the user and intensify the enjoyment of the game, not all elements work well or are appropriate for the work space. This section here will explain some of the most commonly used and recognizable game elements used in “gamification” today.

The use of *achievements* is found in nearly all games made today. Achievements are used to give the user a feeling of accomplishment, motivating the user to continue in order to feel the sense of accomplishment again. In fact, according to Bunchball, a company specializing in “gamification”, achievements can satisfy all of the human desires for work: reward, status, achievement, self-expression, competition and altruism [7]. Reward is self-explanatory. If the user gets an achievement, he or she is normally rewarded with a prize of some type such as points or a bonus [7]. Status is the recognition from others that one gets from accomplishing a certain achievement [7]. Achievement, as mentioned before, is the sense of accomplishment by fulfilling all the tasks required for an achievement [7]. Achievements fulfill the desire for self-expression by being a way to show what one has done recently or is good at. This can be done by viewing the types of achievements one gets as well as when they made them [7]. Competition is the contest between fellow users to accomplish more achievements than the others or be the first to accomplish a certain achievement [7]. Altruism is a form of advertisement for the game. An example would be gifting, giving in-game items to others for free. Many achievements in games today include

achievements such as send gifts to 100 friends or craft a certain amount of gifts. People who receive gifts would feel the urge to go into the game in order to redeem their gift as well as make gifts to send to their friends in order to get the achievements as well.

Another main element used in “gamification” today is *points*. In games, points come in various forms and with various uses. One of the main uses of points in games is to calculate the scores. In many arcade games, players attempt to get as many points as possible in the game as points are directly correlated with how well you do in the game. Points can also be used as a form of currency in games. For example, in standard role-playing games where there players do not receive a score, “gold”, “credits”, “tokens” and “experience” are used instead as a way where you can compare yourself to other players. In such games, players often rate themselves to other players by comparing the rarity of their publicly viewable equipment such as armor and weapons or add-ons such as pets. Many of these items are purchasable with in-game currency meaning that many players would focus on acquiring as many of these “points”. These arguments could also be used for “gamification”. In many games, especially in massively multiplayer role-playing games, players are so focused on obtaining as many “points” as possible to the point where they start using repetitive, boring or non-enjoyable methods to acquire them as fast as possible. An example of this would be “farming”, where the player replays the same level or kills the same monsters in the same area over and over in order to find a certain item or obtain more “points”. This shows that as long as it is for “points” that can help differentiate one user from another or gives a user the feeling that something has been achieved [7], users are willing to do repetitive and often boring actions for long periods of time. This game element would work particularly well in manufacturing lines where the workers often have to install the same part or do the same action over and over. Furthermore, as mentioned in the previous section, the point system is already being used by airlines to good effect [8] and can be used in other areas of business such as cafés or restaurants.

A third popular and effective “gamification” element is *leaderboards*. Leaderboards are scoreboards which show who the top players of a particular game are, often organized by points or levels. This helps foster competition between players in order to raise their rankings or as being on the leaderboards mean that they are one of the best at something. Using this element at work can help motivate workers to constantly improve in order to raise or keep their rankings as well as challenge their fellow users to do better.

Related Work

Over the last decades, “gamification” and more recently computerized cognitive assessment have been increasingly getting more attention in several fields of application. In fact, in recent history, there have been several studies that demonstrate and analyze the effect of computerized

cognitive assessment on people with impairments as well as the elderly and young children. An example of such a study is “Computerised cognitive assessment of athletes with sports related head injury” by A. Collie, D. Darby and P. Maruff from the University of Melbourne, Mental Health Research Institute of Victoria as well as the La Trobe University [11]. This paper analyzes different software-tools such as the Cambridge Neuropsychological Test Automated Battery (CANTAB) by Sahakian et al. and the Automated Neuropsychological Assessment Metrics (ANAM) by Bleiberg et al that could be used for the cognitive assessment of athletes with head injuries from contact sports such as rugby or boxing [11].

Another example of studies that use software for cognitive assessment is “Virtual Reality and Cognitive Assessment and Rehabilitation: The State of the Art” by A. Rizzo and J. Buckwalter from the University of California [12]. This study assessed the advantages and disadvantages of using virtual reality to assess cognitive function of participants with “brain injuries and neurological disorders” as well as analyze the possibility of using virtual reality software to rehabilitate the said participants [12].

STATE OF THE ART

GATRAS

Games to Train and Assess Impaired Persons, more commonly known as GATRAS, is a battery of tests developed by the University of Stuttgart. The main goal of the software is to assess and assist the motoric and cognitive abilities of a person with impairments. It currently consists of eight games: Snakes, Circles, Maze, Pong, Monkey Ladder, Combination, Shapes and Rotations. Of these eight games, Snakes, Circles and Pong solely evaluate the motoric abilities of the participant and were not included in the study.

In each of the games, the scoring worked as follows:

1. Each game starts in level 1
2. If one successfully completes a level without making any mistakes, then the next more complex level is generated and the user is given points equal to the level he or she just completed (i.e. completing level 1 gives 1 point, level 2 gives 2 points, etc.).
3. If level 5 is successfully completed without any mistakes, the user receives 5 points and a new level 5 is generated but the difficulty stays the same.
4. If one makes a mistake during the course of the game, the difficulty is lowered by one and he or she is sent to the beginning of a new semi-random level of the previous difficulty (i.e. if one makes a mistake in level 4 then he or she is immediately sent back to level 3).
5. If one makes a mistake in level 1, then the user has to start over again with a newly generated level 1.

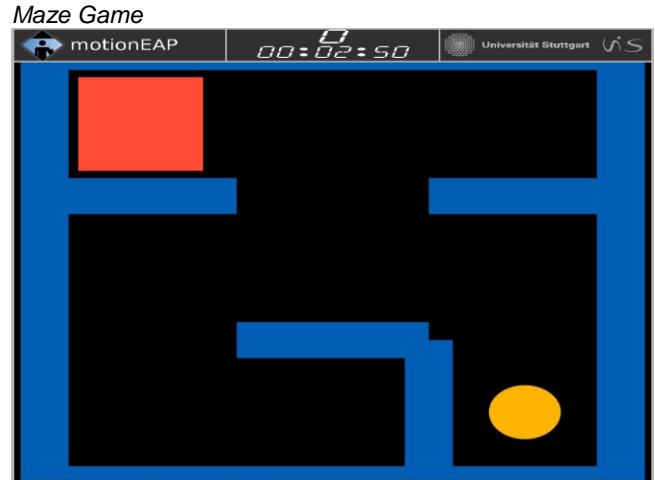


Figure 1. One of the possible level 1 scenarios in the Maze Game.

Originally, the Maze game in GATRAS evaluated both the motoric and cognitive abilities of the participant as it required the user to both quickly and steadily bring the ball through the maze and into the red box as well as find the correct paths in the increasingly difficult and semi-random mazes. However, as the study only evaluates software-tools that assess cognitive abilities and the source code of GATRAS is easily accessible to students at the University of Stuttgart, this game was then slightly modified. In the original version, if one touched any of the blue walls in the maze, then the software automatically recognized the contact as an error and returned the user to the previous level. As this was mainly a motoric assessment concern, this was removed in the modified version of the game. However, the number of times the user touched the walls were still recorded.

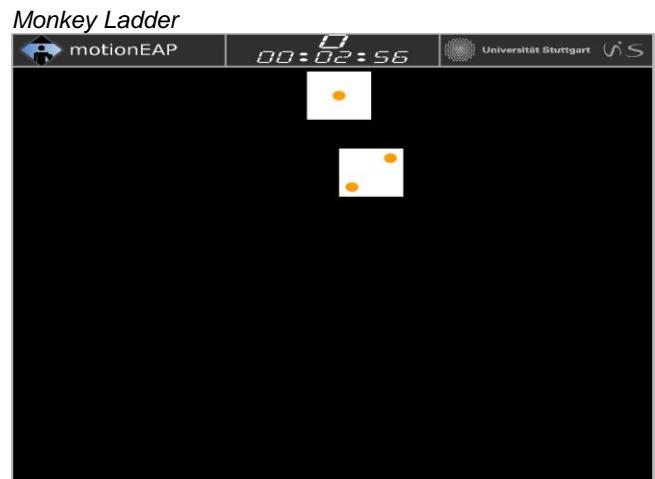


Figure 2. One of the possible level 1 scenarios in the Monkey Ladder Game.

This game assesses the cognitive abilities of the participant as the participant needs to memorize which blocks are

assigned which numbers and select them in numerical order. At the beginning of each level, all the numbered blocks are shown with the corresponding number of dots on it. When the participant first taps on the block with one dot on it, the dots on the other blocks disappear. The participants then have to use their memory to remember how many dots each block has. If the participant manages to click on all the blocks in the correct order, then the level increases by one and the number of blocks increases. If the participant successfully completes the level at level 5, the number of blocks is kept the same but the participant receives 5 points to his or her score. Furthermore, the position of each block is randomly generated so the participant will not be able to complete the same levels with body memory, completing a level through repetition of the same actions. If the participant does not click on the block with only one dot first or clicks the blocks in the wrong order, a mistake will be registered and the number of blocks and level will be decreased by one. However, if the participant is already at level one when he or she makes a mistake, then the mistake is registered but the level stays the same, albeit with a newly generated level.

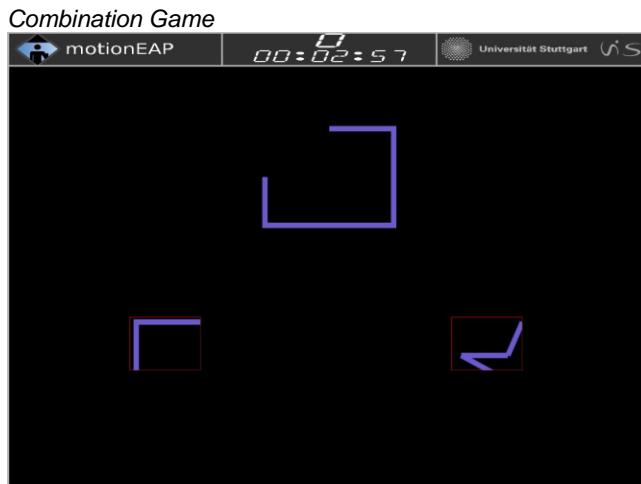


Figure 3. One of the possible level 1 scenarios in the Combination Game.

The Combination Game evaluates the participant's ability to understand symmetry and the correct form of shapes. In this game, one simply needs to choose from one of the up to six possible pictures on the bottom on the screen which can be used to complete the main picture on the upper half of the screen. For example, in **Figure 3**, the user would simply need to click the image on the lower left side of the screen in order to proceed to the next level.

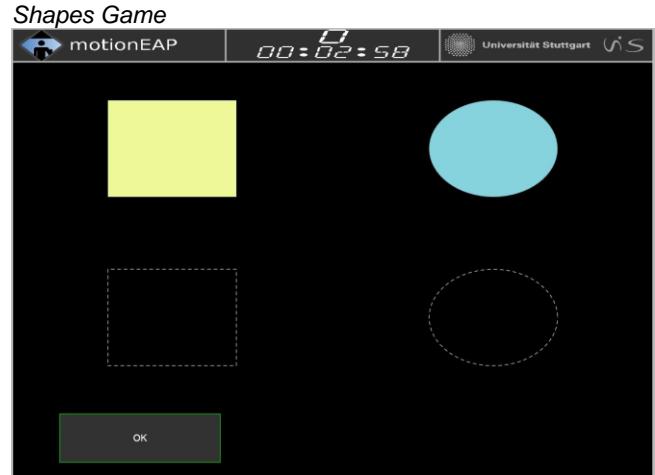


Figure 4. One of the possible level 1 scenarios in the Shapes Game.

The Shapes Game evaluates the participant's cognitive ability to match and recognize similar shapes. In this game, the user needs to match the shapes on the upper half of the screen with the outlines on the bottom half of the screen. After all the shapes are dragged to their corresponding outlines, the user needs to press "OK" in order to proceed to the next level. For example, in **Figure 4**, one needs to drag the square on the upper left corner of the screen to the outline of the square in the bottom left corner of the screen and the circle on the upper right corner of the screen to the outline of the circle on the bottom right corner of the screen before pressing "OK" to proceed to the next level.

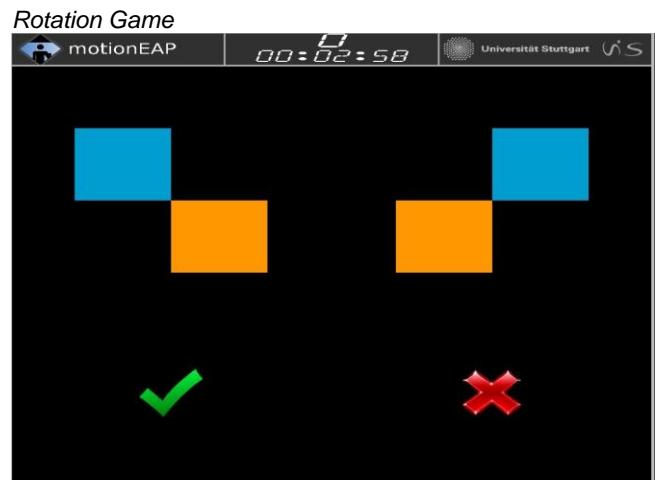


Figure 5. One of the possible level 1 scenarios in the Rotation Game.

The last game in the cognitive section of GATRAS is the Rotation Game. This game assesses the participant's spatial awareness ability as well as the participant's ability to mentally rotate shapes in his or her mind. Here, the user needs to mentally rotate the image on the upper left section of the screen clockwise by 90 degrees and then decide

whether or not the rotated image visualized in his or her mind is the same as the image shown on the upper right section of the screen. For example, in **Figure 5**, the image on the left matches with the image on the right if rotated clockwise by 90 degrees. Therefore, the user would need to click the green checkmark on the lower left section of the screen to proceed to the next level.

CogState

Unless referenced otherwise, all information from this section comes from the CogState Research Manual [13] and CogState Task Descriptions [14] by CogState Limited.

The CogState software from CogState Research comes with 14 games which tests different areas of the participant's cognitive abilities: International Shopping List Task, Chase Test, Groton Maze Learning Test, Fixed Response Mapping Task, Detection Task, Identification Task, One Card Learning Task, One Back Task, Two Back Task, Set-Shifting Task, Continuous Paired Associate learning Task, Social-Emotional Cognition Task, Groton Maze Learning Test – Delayed Recall and International Shopping List Task – Delayed Recall. With these tasks, multiple test batteries can be made to test people with different disabilities such as the pre-made "Early Phase Battery" which tests a broad range of cognitive functions as well as the "ADHD Battery" which is optimized to test people with attention deficit hyperactivity disorders. As all our participants have varying types and degrees of disabilities, the "Early Phase Battery" recommended in the CogState Research manual was used. The "Early Phase Battery" consists of 4 tasks: Groton Maze Learning Task, Detection Task, Identification Task and One Card Learning Task.

Groton Maze Task

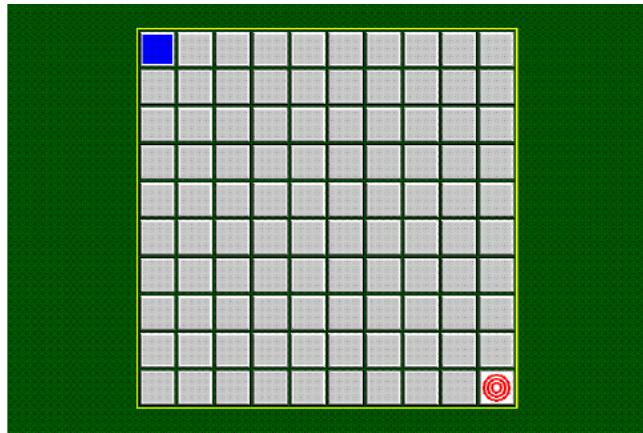


Figure 6. Groton Maze Learning Task.

The main goal of the Groton Maze Task is to test the memory of the participant. In this task, the participant has to find the invisible path from the blue square in **Figure 6** to the red target on the bottom right corner of the grid by clicking the grey squares. In this task, the participant can only move 1 square horizontally or vertically but not

diagonally. If he or she clicks the wrong field, an error sign will be shown. Afterwards, the participant needs to click on the last known correct field. If the participant manages to find the hidden path, which is never fully shown as the participant can only see the previously clicked field marked in blue, the participant needs to repeat the task 4 more times where the participant needs to find the exact same path before proceeding to the next task.

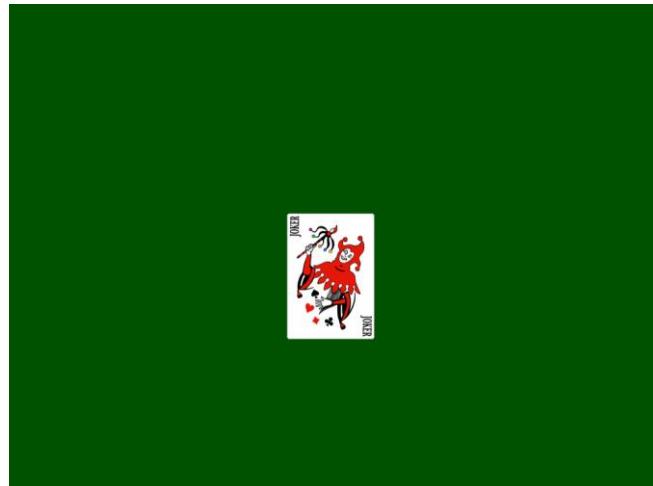


Figure 7. Detection Task.

In this task the participants had their reaction times tested. Here, they had to click "yes", whenever the card in the middle of the screen turned over. While this task sounds quite simple, it is made more difficult (especially for the participants with previous computer experience) because the "yes" button was mapped to the right mouse button even though in most computer programs, the "yes" action would be mapped to the left mouse button. Clicking the wrong button or clicking too early or too late would count as a mistake. For example, if the participant saw the image in **Figure 7**, he or she would have to click the "yes" mouse button (right click) as fast as possible.

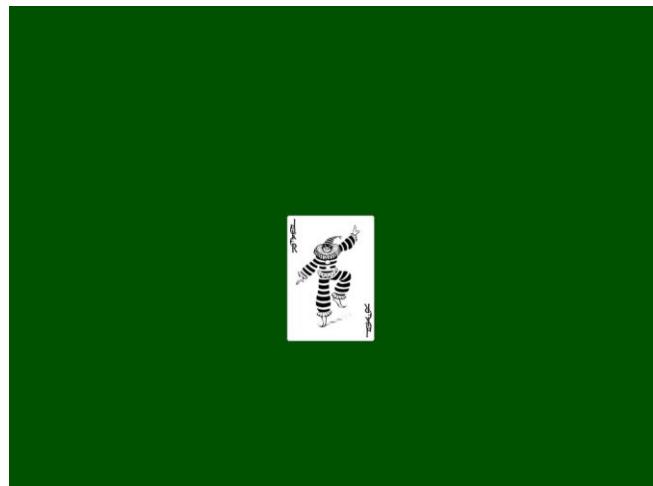


Figure 8. Identification Task.

The Identification Task is an extension of the Detection Task. In this test, the participants had to click “yes” when the card was red and “no” when the card was black. For instance, as soon as the participant sees the image in **Figure 8**, he or she would have to click “no” (left click).

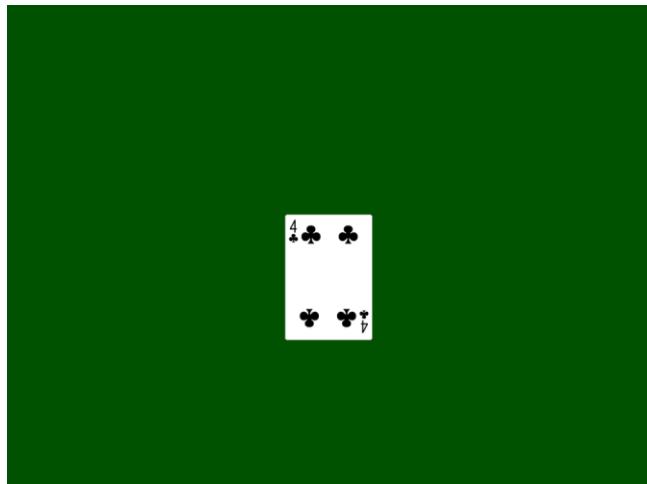


Figure 9. One Card Learning Task.

The final test of the “Early Phase Battery” is the One Card Learning Task. It was a combination of the Detection and Identification Task. The goal of this test is to measure the reaction time and the memory of the participants. Here, the participants needed to remember every card that they saw and click “yes” if they’ve seen it before and “no” if they haven’t seen it yet. In addition, due to the long test time and the sheer amount of cards, this can be seen as one of the most difficult games in CogState. Using **Figure 9** as an example, the participant would have to click “yes” if the four of clubs has appeared at least once before. However, if the participant believes that it is the first time he or she is seeing this card in this test, he or she should click “no”.

hamet e

Unless referenced otherwise, all information from this section is inferred from “Hamet” by Trainsition! [15] and Berufsbildungswerk Waiblingen [16].

The hamet tests are a set of tests which have been developed to assess people with various disabilities in order to suggest occupational fields where they would possibly be most successful in. Currently there are 2 different hamet tests, *hamet 2* and *hamet e*, which are developed for people with specific impairments as well as varying difficulties. Due to the general difficulty of the *hamet 2* tests, Berufsbildungswerk Waiblingen recommended the *hamet e* tests to be used for the study.

The *hamet e* test was developed for assessing the basic motoric and physical abilities of mentally disabled persons. The procedure is based on the *hamet 2* test; however, there are some differences within the tests and assessment methods in order to adapt it to the capabilities of the target group and the environment of a sheltered workshop.

According to the Berufsbildungswerk Waiblingen, the tests assess the same seven factors as the *hamet 2* test, each with four increasing levels of difficulty where the first level represents a simple work applicable tasks and the last level the transition to *hamet 2*:

1. Routine and Speed
2. Simple Tool Insertion and Tool Control
3. Awareness and Understanding of Symmetry
4. Understanding and Application of Instructions
5. Complex Tool Insertion and Tool Control
6. Accuracy of Measurement and Precision
7. Task Specific Factors

The practical execution of the test is usually performed by specially trained instructors or occupational therapists. Furthermore, specialized tools such as a work bench and a bench vice are required to accomplish the tasks.

In order to create a usable result, the *hamet e* test normally requires at least four to five hours for each participant to complete. Because of a participant’s motivation and concentration loss, it is common to split the tests into units of at most two hours per day.

The *hamet e* test can be used to assess the motoric abilities of mentally disabled persons in various areas. Essentially, there are three main fields of application for the test:

1. Use as an entrance test at the beginning of a sheltered workshop
2. Transition into the field of work in a sheltered work environment
3. For creating and assessing integration, aid, and educational schemes

Due to several factors such as time constraints and the extreme variance in types and degrees of cognitive disabilities that our participants had, only the Level 2 tests in the *hamet e* computerized tests were used after discussing with the Berufsbildungswerk Waiblingen.

Data Transferal

Figure 10. The order information displayed on the screen. The participant needs to fill out the missing information in the blue field.

The Level 2 tests of hamet e consists of five tests. The first can be seen in **Figure 10** and is called “Data Transferal”. The participant receives a physical list with a detailed description of the order. His or her job is to type the quantity of the order into the blue field and press “continue” if the information on the list corresponds with the information on the screen. This is a task, which is common for many office jobs. If a wrong amount is entered, it is counted as an error. Fundamentally, this test verifies the participant’s ability to transfer information from the physical world to the PC.

Write SMS

Figure 11. The graphical model of a cellphone and the given instructions.

The next test is the “Write SMS” test. Here, the participant has to write an SMS by pressing the buttons on the graphical model of a cellphone on the screen. To do so he has to follow the instruction given on the right (see **Figure 11**). Pressing a button other than the one mentioned in the

description is counted as an error. This test evaluates the user’s ability to correctly follow the given instructions as well as his or her ability to remember the positions of the buttons on the cellphone in order to reduce the time required to complete the task and the number of errors.

Input Coordinates

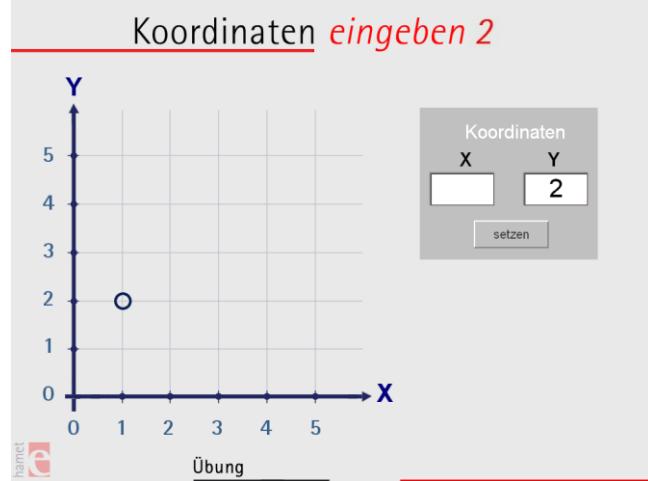


Figure 12. The coordinate system with the given point. In this example, the x-coordinate needs to be given by the participant.

The third test is the “Input Coordinates” task, where the user has to type the coordinates of a given point shown on the graph. In the Level 2 test, either the X- or Y-coordinates will be missing and the participant then has to enter the missing coordinates in the empty white field on the right (see **Figure 12**). After entering the coordinates, the participant needs to press the “set” button in order to continue with the next exercise task. Entering the wrong x- or y-coordinate is counted as an error.

Type Words



Figure 13. Example of the Type Words task.

The second to the last test is the “Type Words” is quite simple. A word is shown in the grey field on the right side of the screen (see **Figure 13**), which must be typed

correctly into the empty white field using the physical keyboard. This is also a common task in office jobs. The participant's ability to memorize and learn the position of the keys on the keyboard is important in this task.

Weigh Package

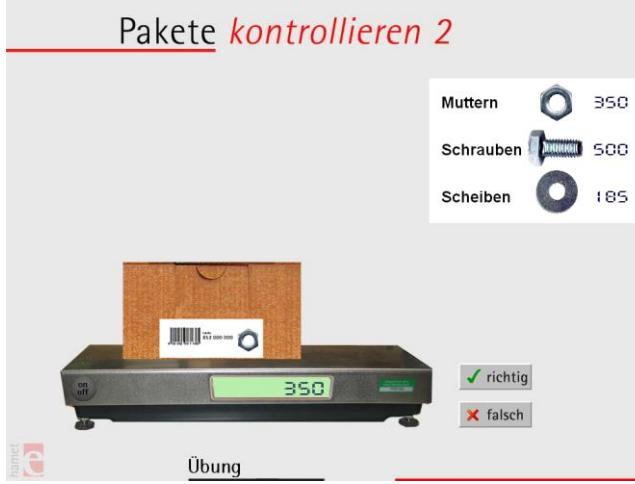


Figure 14. The weight scale with the package and the check list.

Last but not least, we have the “Weigh Package” test which shows a weight scale with a package on top of it. The scale displays the weight of the package and the user will be asked whether the weight shown is correct or not. If the displayed weight is correct, then the participant needs to press the “correct” button, otherwise the button “incorrect” needs to be pressed. In order to determine whether the weight is correct or not, a check list is shown at the right side of the screen (see **Figure 14**). A false answer is counted as an error. For example, a package of nuts needs to have a weight of 350. The scale shows that amount, thus the weight is correct.

Ground Truth Study

The Ground Truth Study was a shear manufacturing test with which the results were then used as a baseline for the three software-tools. Before the test began for each participant, the supervisor explained the manufacturing process and then the participants were allowed to manufacture a shear as practice before moving on to the timed trials. Afterwards, the participants were asked to produce five shears each where the time required to manufacture each shear was recorded.

Shear Manufacturing Process

To manufacture a shear, each of the participants had to follow 9 basic steps.



Figure 15. Manufacturing area for shears.

Step 1: The participant had to sit in front of a table as shown in **Figure 15**. The surface allows the participant to manufacture 3 shears in parallel but for this study, it was decided to take the time it takes to manufacture 5 shears in series.



Figure 16. Example of how the screw should be attached to the production platform.

Step 2: To start the manufacturing process, the participant needed to take a screw from the left most box in the picture on the left and magnetically attach the head to one of the 3 production platforms.



Figure 17. Example of how the screw should be linked through the bottom half of the shears.

Step 3: The participant has to take the right half of the shear from the box on the left side of where he or she is sitting and link the screw through the hole.



Figure 18. Example of how the shears-halves should be placed on top of one another.

Step 4: The participant then has to take the left half of the shear from the box on the right side of the production

platform and put it on top of the right side of the shear as shown in **Figure 18**.



Figure 19. Example of how the washer should properly sit on the shear-halves.

Step 5: The participant then has to link the plain washer (second left-most box in **Figure 19**) through the screw.

He or she also has to make sure that the correct side of the washer (the side without small circles engraved on the washer) is facing upwards.



Figure 20. Example of the correct position of the split-washer.

Step 6: The participant then has to link the split washer (right-most box in Figure 20) through the screw.

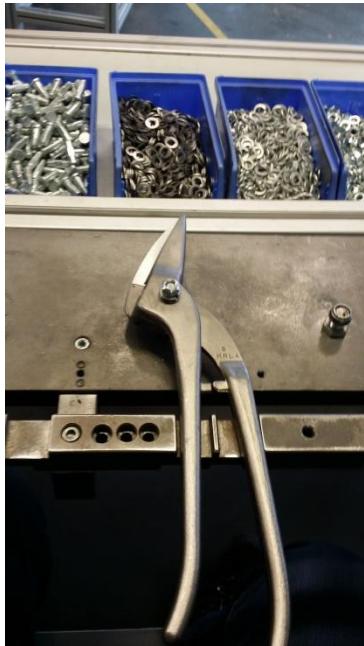


Figure 21. Example of nut position on the screw.

Step 7: The participant then has to screw the nut on the bolt. The majority of the participants had problems with this step as they had problems making sure that the nut was evenly screwed on the bolt.

The participant also has to make sure that the curved side of the nut was facing upwards.



Figure 22. Example of the electronic nut driver in use.

Step 8: The participant then had to take the electronic nut driver and tightly screw the nut on the bolt until he or she heard a beep from the machine.



Figure 23. Example of how the shears should be placed on the belt.

Step 9: Last of all, the participant had to take the completed shear off the production platform and lay it on the belt. Only after this step was completed did the testers stop the stopwatch.

COMPARISON

Requirements

In its current form, the GATRAS software has a few basic system requirements. First of all, it is only fully compatible with systems running Microsoft Windows XP and above. Furthermore, as it was developed for use with a PixelSense Table, it has a minimum resolution of 1600x900. However, when using non-widescreen devices, parts of the lower half of the software are not visible.

Furthermore, the GATRAS software currently needs a minimum of 90 MB hard drive space to install unless it is run from a USB device.

On the other hand, the CogState software is compatible with Microsoft XP and Vista [13]. According to the minimum hardware specifications released by CogState, it requires 10 GB hard drive space, 1024 MB RAM, an internet connection and a minimum resolution of 1024x768 [13].

Finally, hamet e is compatible with all versions of Microsoft Windows. Furthermore, unless it is run from an USB device, it requires a minimum of 600 MB in the hard drive to install.

Usability

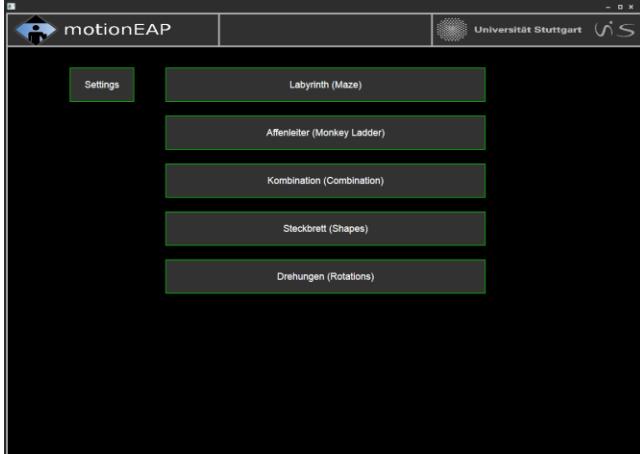


Figure 24. Main Menu in GATRAS software.

As can be seen in **Figure 24**, the main menu in GATRAS is quite simple to use. The button in the upper left corner of the screen allows access to the settings and the buttons in the middle of the screen are the games which can be used for cognitive assessment of the participants.

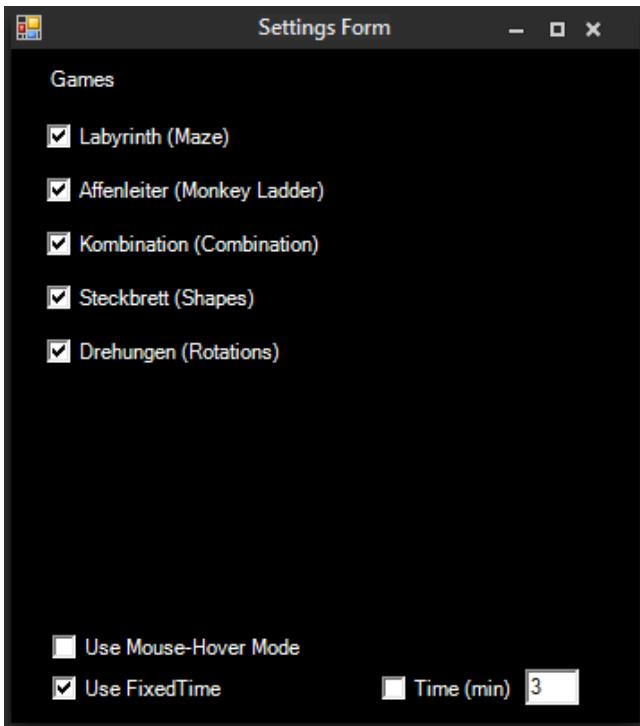


Figure 25. Settings screen in GATRAS

In the settings screen (**Figure 25**) the user can choose which games are available in the main menu. The Mouse-Hover Mode is an option mainly made for use with the touchscreen. With this option selected, one can select an object by dragging his or her finger over it instead of having to tap and hold the object. The settings screen also allows the user to disable and enable the time limit as well as set the number of minutes each game should be played.

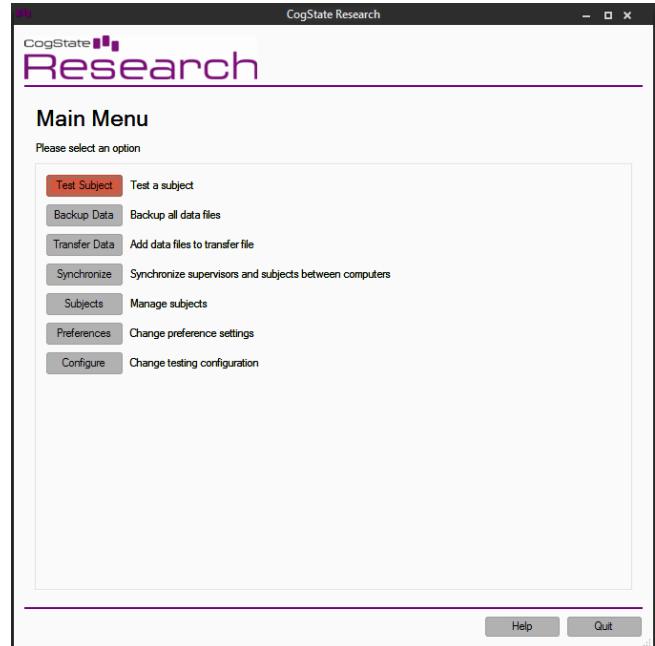


Figure 26. Main Menu in CogState.

Similar to GATRAS, the main menu from CogState is quite easy to understand. Every button in the main menu comes with a description of what it does (i.e. the “Configuration” button allows the user to change the testing configurations). From the main menu, the user can begin to test a participant, backup a local copy of the test data to the computer, transfer data to the DataPoint website for analysis, synchronize the data from test instructors and participants between computers with the same account, add a participant or edit an existing participant, change software options such as language or whether to transfer data via internet or USB device and select the tests in the test battery.

Figure 27. Subject Management screen in CogState.

When the user clicks “Subjects” in the main menu seen in **Figure 26**, he or she will see a similar image to the one in **Figure 27**. Here, the user can add new participants by entering the screening ID of the participant, optionally a randomization ID, gender of the participant, whether the participant is left or right handed, year of birth of the participant as well as the preferred instructing language for the participant. This window also allows the user to edit the information of any participant already saved in the system.

Figure 28. Test Battery selection in CogState

When the user clicks “Configure” in the main menu in **Figure 26**, he or she is asked to enter in the activation code provided by CogState Research. If the activation code is valid, he or she will be sent to the page seen in **Figure 28**. Here one can select and save the tasks in the test battery.

Figure 29. Main Menu of DataPoint website in CogState.

In the main menu of the DataPoint website seen in **Figure 29**, the user can view and see the analyzed test data from the participants. It allows the participants to be sorted by session date, upload date or by name. Furthermore, the user can access material such as the manual and task descriptions in this website.

Figure 30. Main Menu for hamet e.

Figure 31. Test selection screen in hamet e.

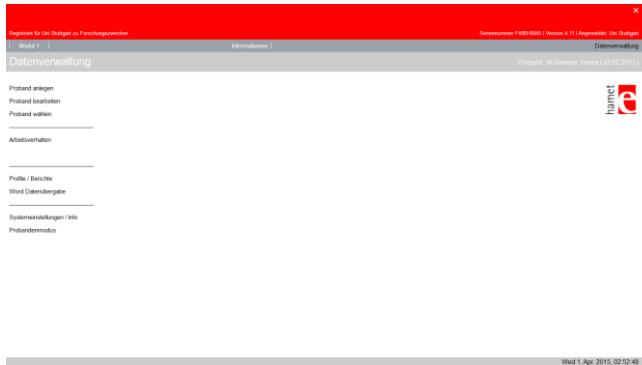


Figure 32. Configuration screen in hamet e.

Unlike the main menus from GATRAS and CogState, the one from hamet e is a little more confusing as all the buttons are widely separated from one another and do not provide a detailed description of what each button does (see **Figure 30**). In the main menu, the user can see the software version, serial number, as well as under which account you are currently logged in as. The button “Modul 1” is used to access the computerized hamet e tests (**Figure 31**), the “Information” button currently serves no function as an empty page opens and the “Datenverwaltung” button leads to the configuration menu (**Figure 32**) where one can add, edit or select a participant, as well as gives access to the data analysis tools.

Figure 33. Adding new participant screen in hamet editor

When the user clicks “Proband anlegen” (add participant) in the configuration menu in **Figure 32**, the user will see the image in **Figure 33**. Here the user has to give in the name, participant ID, birthday and gender of the participant before the user can save the profile. Furthermore, the user can give optional information such as current or desired career of the participant, the type and severity of the participant’s impairment, the participant’s education level, comparison group (whether or not the participant should be grouped with the results of special needs students, secondary school students, high school students or adults). The user is also able to provide other information such as the place where the participant is diagnosed, the attitude of the participant as well as the other comments that the user feels is important to add. Clicking “Proband bearbeiten” (edit participant) in the configuration menu gives the user a similar page to

Figure 32 with the only difference being that some of the fields have already been filled.

Modability

With the GATRAS and hamet e softwares, the test instructor simply needs to select the tests or tells the participants which tests the participants need to do. The values for the results of any tests which are not used will simply show up as "0" for GATRAS and empty for hamet e. Furthermore, the tests can be done in any order. With CogState, the test instructor needs to select the tasks in the test battery before the testing can start. However, each the task selected in the test battery needs to be completed to proceed to the next task and the order of tasks cannot be changed.

Portability

Both hamet e and GATRAS do not currently require any installation and are executable from USB devices. However, CogState requires a relatively lengthy installation and activation process before the software can be used.

Security

Figure 34. Login screen for hamet e.

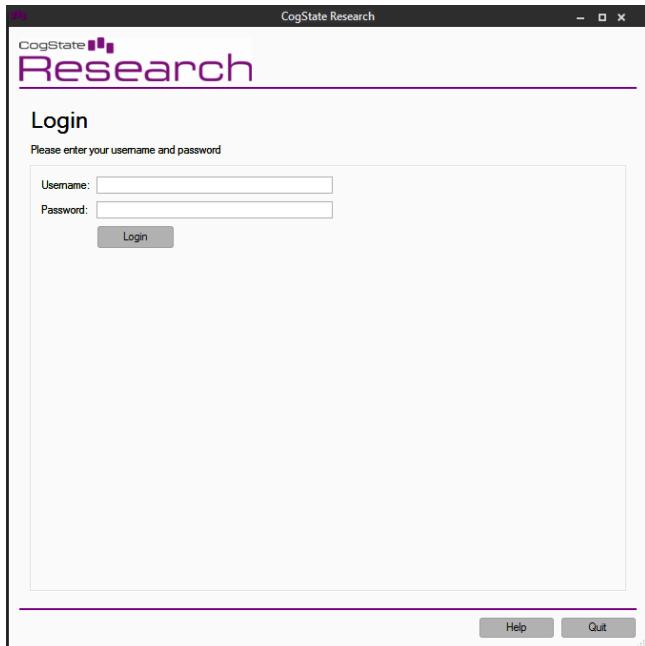


Figure 35. Login screen in CogState.



Figure 36. Login screen for DataPoint website for CogState.

Both hamet e (Figure 34) and CogState (Figure 35) requires the test instructor to log into the software before the user can access the functions of the software such as executing tests as well as editing and creating participants. Furthermore, test data from hamet e can only be accessed from within the software and the test data from CogState can only be accessed through DataPoint after a login process (Figure 36). On the other hand, GATRAS currently has no security functions. There is no login required or available to access the software and the test data can be accessed by opening the software files.

Functionality

While each of the described tests mentioned above evaluates and assesses various cognitive functions of the participant, each of them is very different from one another and therefore, has different advantages and disadvantages in every situation.

Feature	GATRAS	CogState	hamet e
Data Analysis	✓	✓	✓
Easy Synchronization with multiple PCs		✓	
Graphing Function		~	
Improvement Tracking	✓		✓
Log-In Function		✓	✓
Multiple Languages		✓	
Pause Function	✓		
Participant Registration		✓	✓
Practice Function	✓	✓	✓
Usable with Touchscreen	✓	~	
Time Limits	✓	~	

Table 1. General overview of features available to the three software-tools.

GATRAS

Advantages

While GATRAS is still in its development stage, it comes with many functions which help supervisors assess the cognitive abilities of people with impairments. One of its best features, which the other two software-tools lack, is the ability to set time limits for the tests. Since the tests from other software-tools do not have time limits, some participants with more severe types and degrees of cognitive disabilities can require more than an hour to complete a single test. This can be problematic with companies, testers as well as participants which have schedules and other appointments to keep as the participants can take between 20 minutes and 4 hours to complete CogState or hamet e. However, with GATRAS, if the testers set the time limit to 3 minutes per test, the testers can determine that the participant will finish the test within 25 minutes including practice and explanation times.

Another feature that GATRAS has that CogState and hamet e lack is the ability to pause during the course of the test. A pause function allows the participant to take a break when necessary. This is advantageous as many participants with cognitive impairments such as ADHD have short attention spans. Furthermore, many participants may also have motoric disabilities who find the tests physically exhausting. In addition, due to the length of the tests, the participants also have the option to pause the test in order to take refreshments or go to the restroom when necessary.

A third important feature is the ability for the participant to test each game before the actual test starts. This allows the participant to practice each game before the actual recorded tests start so that the tester can determine whether or not the participants have understood the goal and function of each test and whether additional instructions are needed.

Another advantageous feature that GATRAS has is the multitude of data analysis tools it has at its disposal. The

Score Analyzer tool bundled with the GATRAS software allows the testers to convert the actions of the participants in the tests into multiple different scores using various scoring methods. Seven scoring methods are available in the Score Analyzer.

1. **Add Level:** The participant gets points equivalent to the level which he or she successfully completes. For example if the participant completes level 1, then he or she will get 1 point or if he or she completes level 4 in a test, 4 points would be given.
2. **Repeating Levels:** The participant only receives point if he or she moves to a higher level. For example, if the participant successfully completes level 1 for the first time, he or she would receive 1 point. However, if he or she makes a mistake in level 2 and then completes level 1 again for the second time, he or she would receive 0 points as he or she had already completed level 1 once before.
3. **Maximum Levels:** Each time the participant completes a level that he or she has completed before, the participant will only receive half the points he or she received the last time the level was completed. For example, if the participant completes level 5 for the first time, he or she would receive 5 points. The second time the participant completes level 5, he or she would only receive 3 points. On the third iteration, he or she would get 2 points, etc.
4. **Repeating Maximum Levels:** Combination of Repeating Levels and Maximum Levels scoring methods.
5. **Add Time Left:** Time left over after completing level 5 for the first time added to the final score of the participant.
6. **Add Time Left with Shift:** Similar to the Add Time Left scoring method. However, every 5 seconds points equal to the current level divided by 2 is deducted from the participant's score.
7. **Scaling Points:** This method rewards players with higher endurance by providing more points for each level for every passed minute of playing a game.

This wide variety of scoring methods is useful for analyzing the results of the participants as well as comparing the results to those from other assessment tools.

Another advantage of the GATRAS tests, are that they are extremely gamified to the point where they are more similar to games than tests. This is very advantageous as playing games is much more fun than doing tests. This helps keep the motivation of the participants high. Keeping the motivation of the participants high is critical as it is important for the mental and physical condition of the participant to stay stable so that it does not influence the

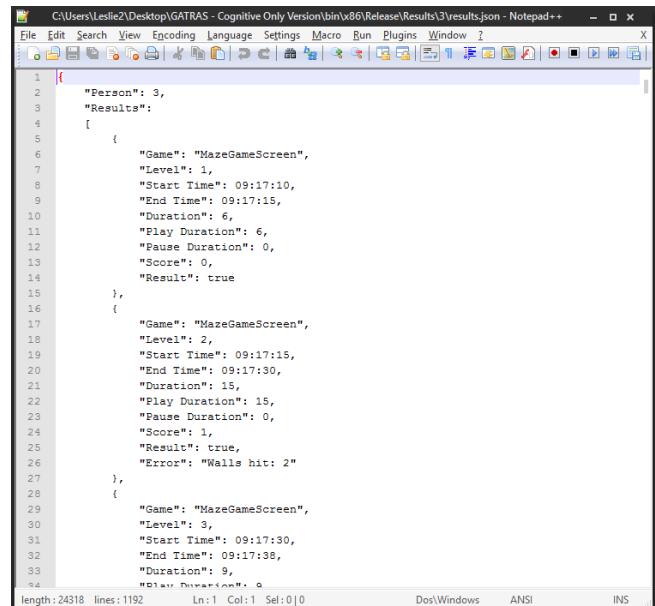
final score in a major way. Furthermore, if a participant had fun doing a test, he or she would be more interested in doing the same test again in the future.

Lastly, GATRAS and its games do not require any prior computer or technical knowledge to use. As it was primarily developed for touchscreen devices at the beginning, all the tests could be done with the mouse and all the buttons and controls were well sized and easily accessible even for participants using a computer for the first time.

Disadvantages

Despite the advantages of the GATRAS tests, GATRAS does come with a few disadvantages. One disadvantage that GATRAS has is that it has very few tests in comparison with the other software-tools. Whereas CogState has 14 tests which assess different areas of cognitive function and hamet e has 20, GATRAS only has 5. While quantity is not a measure for quality, having a larger amount of tests is advantageous as it allows the tester to tailor the test battery to his or her needs (for example, a mix of tests with a focus on memorization or concentration).

Another disadvantage of GATRAS is that it does not yet allow the testers to register the participants in the software. This makes it difficult for the testers and supervisors to organize the results as the tester needs to manually write down or remember the order in which the participants took the test to match the results with the participants. The lack of participant registration also reduces the readability of the results as the software just assigns a generic number based on the number of participants that used the test before him or her. In addition, the tester is required to exit and restart the program after every participant so that the software knows that there is a new user.

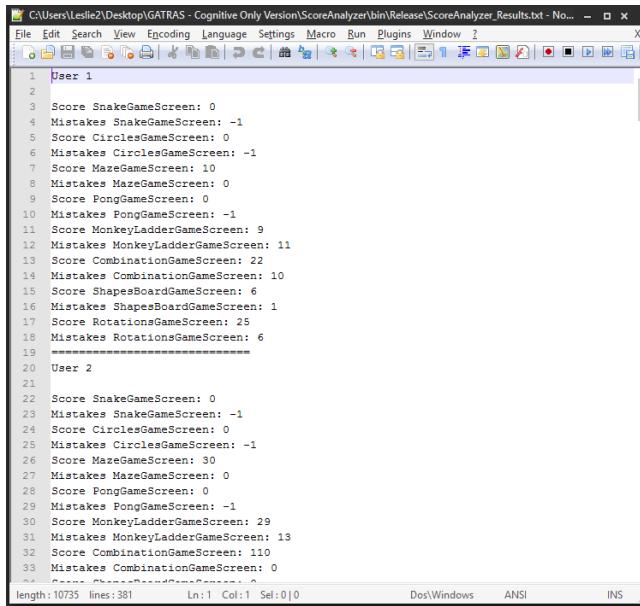


```

C:\Users\Leslie2\Desktop\GATRAS - Cognitive Only Version\bin\x86\Release\Results\3\results.json - Notepad++ - x
File Edit Search View Encoding Language Settings Macro Run Plugins Window ? x
File Edit Search View Encoding Language Settings Macro Run Plugins Window ? x
1
2   "Person": 3,
3   "Results": [
4     {
5       "Game": "MazeGameScreen",
6       "Level": 1,
7       "Start Time": "09:17:10",
8       "End Time": "09:17:15",
9       "Duration": 6,
10      "Play Duration": 6,
11      "Pause Duration": 0,
12      "Score": 0,
13      "Result": true
14    },
15    {
16      "Game": "MazeGameScreen",
17      "Level": 2,
18      "Start Time": "09:17:15",
19      "End Time": "09:17:30",
20      "Duration": 15,
21      "Play Duration": 15,
22      "Pause Duration": 0,
23      "Score": 1,
24      "Result": true,
25      "Error": "Walls hit: 2"
26    },
27    {
28      "Game": "MazeGameScreen",
29      "Level": 3,
30      "Start Time": "09:17:30",
31      "End Time": "09:17:38",
32      "Duration": 8,
33      "Play Duration": 8
34    }
35
36 length: 24318 lines:1192 Ln:1 Col:1 Sel:0|0 DosWindows ANSI INS

```

Figure 37. Example of results from GATRAS. Everytime the participant makes and error or completes a level is recorded in this document. Everytime the participant makes a mistake or completes a level, the software records: the name of the game, the level they completed or made the error in, the starting time of the level, the time it takes to complete the level or make a mistake, the duration of the level, current score and the result of the action (true = level successfully completed, else error message such as “Walls Hit” for the Maze Game.



```

1 User 1
2
3 Score SnakeGameScreen: 0
4 Mistakes SnakeGameScreen: -1
5 Score CirclesGameScreen: 0
6 Mistakes CirclesGameScreen: -1
7 Score MazeGameScreen: 10
8 Mistakes MazeGameScreen: 0
9 Score PongGameScreen: 0
10 Mistakes PongGameScreen: 0
11 Score MonkeyLadderGameScreen: 9
12 Mistakes MonkeyLadderGameScreen: 1
13 Score CombinationGameScreen: 11
14 Mistakes CombinationGameScreen: 22
15 Score ShapesBoardGameScreen: 6
16 Mistakes ShapesBoardGameScreen: 1
17 Score RotationsGameScreen: 25
18 Mistakes RotationsGameScreen: 6
19 =====
20 User 2
21
22 Score SnakeGameScreen: 0
23 Mistakes SnakeGameScreen: -1
24 Score CirclesGameScreen: 0
25 Mistakes CirclesGameScreen: -1
26 Score MazeGameScreen: 30
27 Mistakes MazeGameScreen: 0
28 Score PongGameScreen: 0
29 Mistakes PongGameScreen: -1
30 Score MonkeyLadderGameScreen: 29
31 Mistakes MonkeyLadderGameScreen: 13
32 Score CombinationGameScreen: 110
33 Mistakes CombinationGameScreen: 0
34 Score ShapesBoardGameScreen: 0
35 Mistakes ShapesBoardGameScreen: 0
36 Score RotationsGameScreen: 0
37 Mistakes RotationsGameScreen: 0
38 =====
length: 10735 lines: 381 Ln:1 Col:1 Sel:0|0 Dos/Windows ANSI INS

```

Figure 38. Example of Score Analyzer results. For every participant, the final score and number of mistakes for each game are shown.

Last of all, as seen in **Figures 37 and 38**, GATRAS only saves the user data as a TXT-document. While the majority of operating systems allow the user to read TXT-documents without having to install any additional software, reading and analyzing the data is difficult especially with large numbers of participants. Furthermore, its inability to save in Excel format (.XLS or .XLSX) makes it difficult for the test instructor to analyze the data in an organized manner.

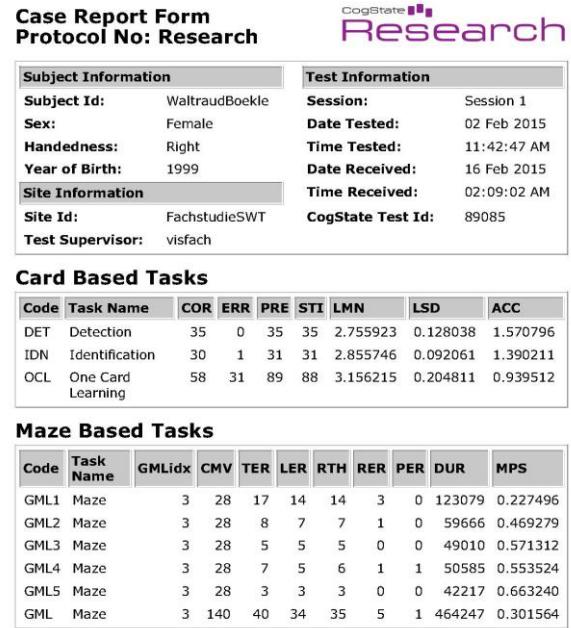
CogState

Advantages

The CogState software has many advantageous features and can be seen as the software with the most features as well as the best well-made software of the three software compared in the study. One of the features that show off its quality are its manuals. The CogState software comes bundled with three documents, a general manual, task descriptions and analysis guidelines. The general manual gives detailed descriptions of the requirements as well as comprehensive instructions on how to install, activate and set up the software. The manual also instructs the reader how to

prepare the test environment before the test as well as instructions on how to upload and use the test data to the DataPoint website by CogState Research. In addition, the manual also offers the reader different test batteries which one can use depending on the cognitive disability of the participant. The task description, as the name states, gives detailed step-by-step instructions of how to proceed in each task as well as giving examples on how to proceed or fail a task. Last of all, analysis guidelines describes “the statistical methods and analyses that applied on data collected from the CogState library” [17]. Basically, it describes how the scores of each task are calculated in addition to hints on how to understand the raw data from the tests.

Another beneficial feature of CogState is that it has a wealth of analysis tools through its DataPoint website. This website allows the user to sort through the participants by name, test date or upload date. Furthermore, the website offers the user the ability to view the test data as a case report form, test report as well as text form with data extraction.



Subject Information								Test Information			
Subject Id:	WaltraudBoekle							Session:	Session 1		
Sex:	Female							Date Tested:	02 Feb 2015		
Handedness:	Right							Time Tested:	11:42:47 AM		
Year of Birth:	1999							Date Received:	16 Feb 2015		
Site Information								Time Received:	02:09:02 AM		
Site Id:	FachstudieSWT							CogState Test Id:	89085		
Test Supervisor:	visfach										

Code	Task Name	COR	ERR	PRE	STI	LMN	LSD	ACC
DET	Detection	35	0	35	35	2.755923	0.128038	1.570796
IDN	Identification	30	1	31	31	2.855746	0.092061	1.390211
OCL	One Card Learning	58	31	89	88	3.156215	0.204811	0.939512

Code	Task Name	GMLidx	CMV	TER	LER	RTH	RER	PER	DUR	MPS
GML1	Maze	3	28	17	14	14	3	0	123079	0.227496
GML2	Maze	3	28	8	7	7	1	0	59666	0.469279
GML3	Maze	3	28	5	5	5	0	0	49010	0.571312
GML4	Maze	3	28	7	5	6	1	1	50585	0.553524
GML5	Maze	3	28	3	3	3	0	0	42217	0.663240
GML	Maze	3	140	40	34	35	5	1	464247	0.301564

Note: If no data was recorded for a task in the CogState battery it will not appear in the above table/s.

Figure 39. Example of Case Report Form from CogState.

The case report form (see **Figure 39**) shows some of the raw data from the tasks. For example, for card based games

such as the Detection Task, it shows the number of correct clicks, incorrect clicks, the number of clicks in total, stimuli, speed of performance, consistency of performance and accuracy of performance. With maze based games such as the Groton Maze Learning Task, the case report form shows the number of correct moves, number of errors, number of legal errors (clicking the same wrong field multiple times), number of times one had to re-click the blue field, number of times the participant did not click an adjacent field or clicked a diagonal field, time required to find the path and number of moves per second for each iteration of the task.

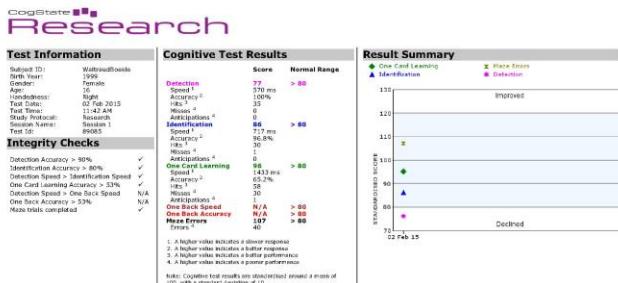


Figure 40. Example of test report from CogState. The left-most column shows the information from the participant (name, age, test date, etc.) and whether or not the participant passed the integrity checks (whether or not the participant completed each test as expected [13]). In the middle column are the cognitive test results from the battery of tests and on the right-most column is a graph detailing any improvements or diminishments in the results over multiple iterations of the test.

The test report (see **Figure 40**) shows the scores for each task, accuracy, number of correct and incorrect clicks and number of premature clicks for card based games and the number of errors for maze based games. Furthermore, the test report has improvement tracking that shows the progress of the participant over multiple iterations of the tests in a graph making it very easy for the user to understand.

Another useful feature of CogState is that requires the user to log in. This is beneficial as it allows multiple studies to use the same CogState software on a single computer. Furthermore, another advantage of having user accounts is that it makes it very easy for the user to gather all the data from a study into a single location by accessing the account in the DataPoint website to view all the results, participants and test dates.

Additionally, the CogState software comes in multiple languages. This is advantageous as it allows the program to

be used across multiple countries as well as saving the test instructors from having to translate the instructions into the participant's native language

Last of all, the CogState software shows the instructions of each task before the participant begins it. This is very beneficial as the test instructors do not need to prepare detailed instructions beforehand removing the risk of improper or false instructions. Furthermore, these written instructions make sure that every participant gets the exact same set of instructions so that no participant has an advantage over another.

Disadvantages

Despite the many positive features of the CogState software, there are quite a few disadvantages as well. One of the main disadvantages is the extreme difficulty of some of the tests for participants with cognitive impairments. An example of this is the Groton Maze Learning Task and the One Card Learning Task. For the Groton Maze Learning Task, finding the hidden path way for the first time was very difficult for the participants as many participants kept forgetting where the last correct position was. Furthermore, the constant mistakes lowered that motivation for many of the participants as they were not receiving any positive feedback from the software. For the One Card Learning Task, other than the difficulty, the length of the test was also too long for many of the participants. After the first 20 cards or so many of the participants started losing interest and just started clicking “no” or “yes” to everything to finish the test faster.

Another disadvantage is that it is sometimes difficult for the participant to determine whether they have made a mistake or not. For tasks such as the Detection, Identification and One Card Learning Tasks, the only feedback that the participant receives when he or she makes a mistake is that the card shuffles to the left instead of the right and the sound is slightly different. This can be seen where some participants marked in the questionnaire that they believed they did not make any mistakes in the CogState tests despite making quite a lot in the test itself. This is seen as a disadvantage as many participants do not realize that they are doing something wrong and continue by making the same mistakes over and over again.

While having user accounts are beneficial in the long run, it also comes with some slight disadvantages. One disadvantage is that it does not come with any option to remember the user name of the account. While it is only a minor annoyance compared to the advantages it brings, it can prove bothersome for accounts with long usernames or studies that need to log in frequently. Furthermore, an internet connection is required. Without an internet connection, one would not be able to upload and download test data to and from the DataPoint website in order to analyze any data obtained in the study.

hamet e

Advantages

Hamet e offers many useful features and advantages over the other two software-tools. The biggest advantage is hamet e's ability to offer the right difficulty level for the right group of users. Unlike with the CogState software where there were tasks that were much too difficult for the user but lacked an option to lower the difficulty, the hamet e software offers 4 difficulty levels for each task. A high difficulty level can easily cause the participant's motivation to drop as well as also cost a lot of time for both the participant and the test instructors. With hamet e this is unlikely to happen as long as the correct level is chosen.

Another advantage of hamet e is that its tests are based on office tasks done in companies everyday such as inputting information into a computer or typing. This is an advantage as it trains the user to do office-oriented tasks while evaluating the cognitive ability of the participant at the same time.

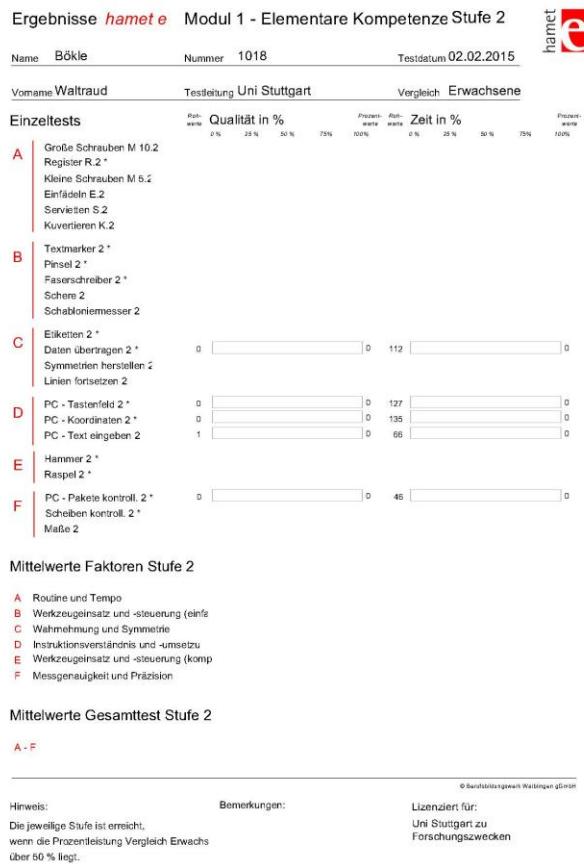


Figure 41. Example of test data from hamet e. For each test, the value on the left side of the left bar symbolizes the number of mistakes made during the test. The value on the right side of the left bar represents the improvement in the number of mistakes compared to previous iterations in percent. The left value on the right

bar signifies the time required to complete the task in seconds and the right value on the right bar is the improvement in time compared to previous iterations in percent.

In addition, the data analysis tools included in the hamet e software also comes with improvement tracking (see **Figure 41**), though only to a certain extent. As mentioned in the CogState Advantages section, improvement tracking is the ability of the software to automatically compare the results of a user to his or her previous results if the participant has completed the test more than once. This is advantageous as the test instructors can see which activity the participant specializes in as well as see their overall progress from any training they might have done.

Furthermore, though not as detailed as the instructions from CogState, hamet e also has written instructions for the test instructors to read to the participants. Additionally, there are instructions in many of the tasks for the participants so that they do not need to memorize every single step of a multi-step task such as Write SMS to complete it.

Disadvantages

As mentioned previously, an advantage of hamet e was that its tests were based on tasks done in offices. However, this can also be seen as a disadvantage as participants that have already had experience working in an office or with computers in general would likely get a better score than someone using computers or doing such tasks for the first time irrespective of cognitive ability.

Furthermore, the data analysis tools from hamet e only tracks improvement to a certain extent as only the average times and number of mistakes from the previous iterations are shown instead of the score from every individual iteration in order to clearly see the progress in numerical or graphical form. Furthermore, unlike GATRAS, the data analysis tools from hamet e allow the user to save a participant's data in several formats such as PDF or LL. However, it should be noted that no Excel (.XLS or .XLSX) formats are available meaning that the score analyst still needs to manually transfer the data in order to analyze it.

Lastly, unlike GATRAS and CogState, hamet e does not come with a scoring system. The software only records the time needed to complete the task and the number of mistakes made while performing the task. This makes it difficult to analyze the data as the user needs to create a scoring system that takes the time and number of mistakes into context. Furthermore, the lack of a scoring system makes it difficult to use results from other studies as there is no standardized score and baseline to compare to without having to recalculate the scores manually.

PRE-STUDY

The pre-study conducted with students from the University of Stuttgart had several goals. The main goal of the pre-study was receive feedback from users in order to test the

likeability, understandability and difficulty of the tests from the software-tools. The secondary goal was to determine whether or not the participants that are relatively unimpaired in comparison with the participants from the main study would perform equally in each software. The pre-study consisted of 5 male and 5 female students from the University of Stuttgart. Before this study, none of the participants had any contact with any of the three software-tools and also did not receive any information regarding the types of tests that they would be evaluated on. Furthermore, all participants were fluent in either German or English in order to ensure that all instructions as well as the questionnaire (**Appendix Figure A1**) were fully understood.

Tools

For the study, three laptops running Windows 7 and 8 were used. On all three laptops, the numpad was disabled and blocked from the participant. The only input tools the participants could rely on were the mouse and laptop keyboard. All three mice used in the tests were roughly the same size and each had only 3 buttons: left, middle and right so that it would not be confusing or complex for an user with little experience with computers.

Procedure

The procedure that each participant in the pre-study has to follow can be split into 4 basic steps.

Step 1: Fill out questions 1 to 3 on the questionnaire. These questions attempt to assess the general condition and experience with cognitive tests or games such as IQ-Tests or puzzles of the participant before they begin with the software tests.

Step 2: Do the GATRAS cognitive software tests. For simplicity's sake, all participants did the GATRAS tests in the same order: Maze Game, Monkey Ladder, Combination Game, Shapes Game and lastly Rotations Game. Before starting each game, the examiner gave a verbal explanation of how the game worked. Afterwards, the participants were given 30 seconds to test out the game in "Test Mode" without the scores being recorded. However, in the pre-study, the majority of the participants opted to start the main test before their 30 second testing period was over. After all tests in GATRAS have been completed, the participants are asked to fill out questions 4 to 8 in the questionnaire. These questions assess the participant's experience with the software-tool such as whether they enjoyed doing the GATRAS games or why they thought they made mistakes during the test.

Step 3: Do the CogState "Early Phase Battery". The order of the test is unchangeable so the participants always do the test in the same order: Groton Maze Learning Task, Detection Task, Identification Task then One Card Learning Task. Before starting each task, the instructor either read the instructions for the task provided by CogState out loud

or let the participant read it him or herself. Afterwards, they were asked to do the exercise tasks provided by CogState before continuing to the recorded session. After the "Early Phase Battery" has been completed, the participants are asked to fill out questions 9 to 13 in the questionnaire which are exactly the same as questions 4 to 8 but with the aim to assess the participants experience with the CogState software.

Step 4: Do the Level 2 tests in the hamet e computerized tests. The software allows the user to do the test in any order but for simplicity's sake, all the participants did the tests in the same order: Data Transferal, Write SMS, Input Coordinates, Type Words and lastly Weigh Package. Similar to CogState, the test instructor either read the instructions on the screen out loud or allowed the participant to read it by themselves. Afterwards, they were asked to complete the unrecorded exercise tasks before continuing to the recorded sessions. After the hamet e computerized tests have been completed, the participants were then asked to complete questions 14 to 18 in the questionnaire.

Results

In this section, the results from the pre-study are analyzed. Using the questionnaire, we attempt to determine the likeability, understandability and difficulty of the tests used in the software-tools. Using the results from the tests themselves, we attempt to fulfill the secondary goal of determining whether or not the participants performed equally in each test.

Questionnaire Results

The objective of the pre-test questions was to evaluate the morale of the participant, as well as find any external factors that could influence the final scores of the participants.

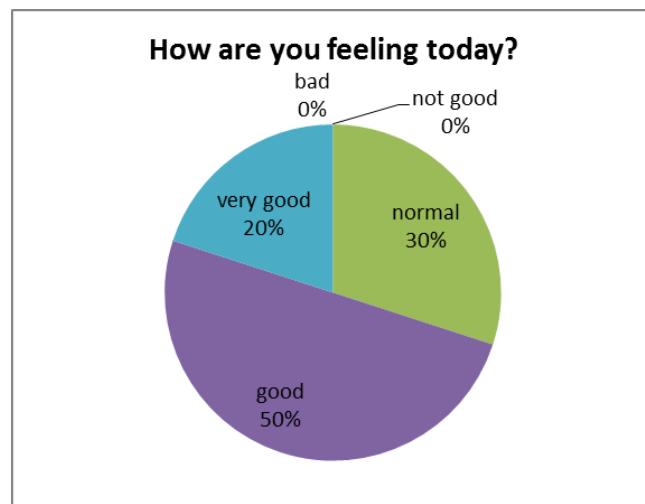


Figure 42. Results from Question 1 from the pre-study questionnaire.

From **Figure 42**, all the participants were feeling normal or better than normal with 70% of the participants feeling “good” or “very good”. This suggest that this group of participants might have a higher than average score as their heightened motivational levels may influence their concentration as well as their motoric and cognitive abilities.

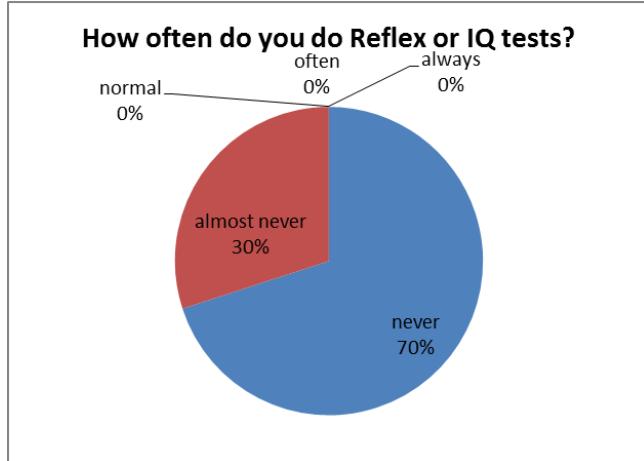


Figure 43. Results from Question 2 from the pre-study questionnaire.

Furthermore 7 of the 10 participants in **Figure 43** had never done a reflex or IQ test before with the other 3 participants only having done an IQ test once before. Since the participants have little to no experience with such tests, the time it takes for the participants to understand and learn the games could potentially be hindered. In the long term, this factor could potentially affect the final scores.



Figure 44. Results from Question 3 from the pre-study questionnaire.

From **Figure 44**, one can see that 60% of the participants played puzzle games at a regular basis. This suggests that they are used to playing games that uses their cognitive skills to solve problems or complete tasks.

Furthermore, as mentioned in **Pre-Study Procedure**, after a participant completes the tests in a software-tool, he or she has to fill out a section of the questionnaire that assesses the difficulty of each software as well as whether they found the software enjoyable enough to do again in the future.

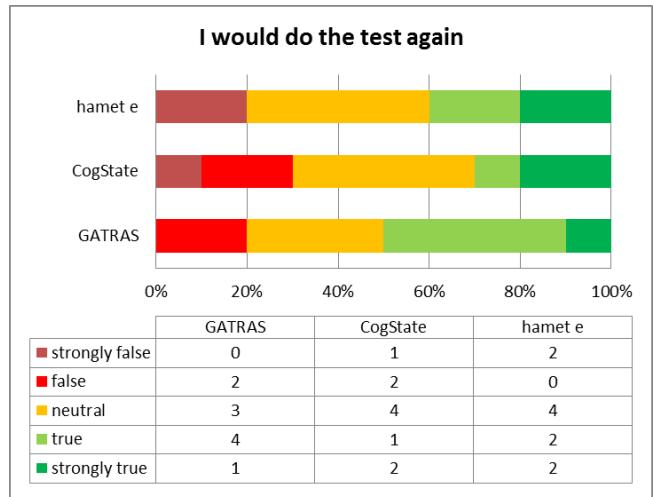


Figure 45. Results from Question 6, 11, 16 from the pre-study questionnaire.

In **Figure 45**, one can see that the participants liked the GATRAS and hamet e softwares the most. 5 of the 10 students believed that they would like to do the GATRAS test sometime in the future again whereas 40% of the participants thought the same for hamet e. On the other hand, only 3 of the 10 participants would like to do CogState again.

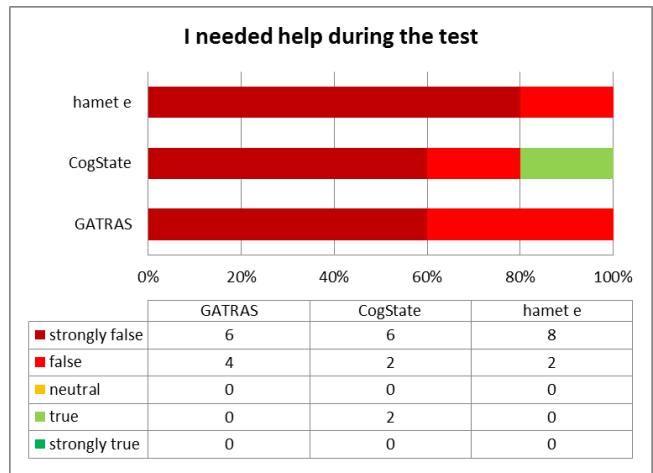


Figure 46. Results from Question 8, 13 and 18 from the pre-study questionnaire.

As can be seen in **Figure 46**, the participants found the tests for all three software-tools easy to understand and do. None of the students said they did not require any help to

understand or learn the tests from GATRAS and hamet e and only 2 of the 10 students needed help to understand and complete the tests in CogState.

In addition, the participants of the pre-study were asked for suggestions for what could be changed in future tests as well as why they thought they made mistakes.

For GATRAS, one of the main complaints was the lack in variety of shapes in the Combination and Shapes Games as there were only 6 different shapes. After completing level 5 of those two games, it became uninteresting for the participants as they always saw the same 6 shapes again and again.

As for CogState, 40% of the participants suggested making the Groton Maze Learning Task easier as there was a wide discrepancy in the difficulty compared to other tests. In addition, 70% of the participants suggested making the One Card Learning Task shorter as many got bored before they completed the task.

For hamet e, the main complaint was that the tests were too easy. However, it should be noted that the majority of students were studying computer science and software engineering and therefore have a lot of experience with computers making the level 2 tests of the hamet e computerized tests trivial for them

Results Analysis

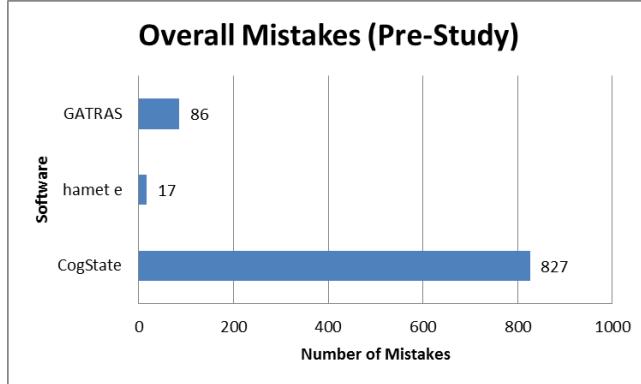


Figure 47. Total number of mistakes made by the participants for each software in the pre-study.

Figure 47 emphasizes the difficulty of the CogState tests by showing the number of mistakes made in total for each software. Over the course of the pre-study, the participants made 827 mistakes in total in CogState. That is 89% of all mistakes made in the pre-study. In comparison, the participants only made 86 mistakes in the GATRAS tests and only 17 in the hamet e tests. To put it in perspective, the participants made nearly 50 times more mistakes in the CogState tests than with the hamet e computerized tests.

Due to the different types of scoring methods used in each test, normalization, the adjustment of values on different scales to a theoretically common scale, was used in order to compare the graphs and data.

Furthermore, since the hamet e and Ground Truth Study did not produce any scores but times instead, the scores were converted via the formula:

$$score = \lceil \max(time(all)) \rceil - (time(participant ID)) / 10 \rceil$$

Formula 1: Equation for computing scores for hamet e and Ground Truth Study results based on the time required to complete the tasks.

Note that the division by 10 was not used for the calculating the “score” for the “Package Control” test for hamet e due to scaling variances. Furthermore, this time to score formula was used so that high scores were better than low scores similar to the scoring methods used in CogState and GATRAS.

For normalization we used the formula:

$$(Score of Participant / Mean Score of Test) * 100$$

Formula 2. Equation for computing normalized values.

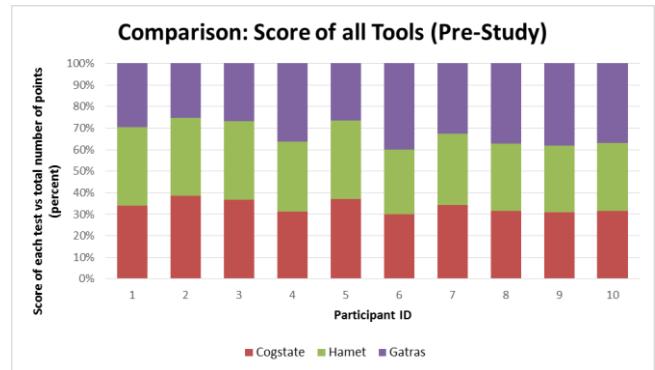


Figure 48. Normalized pre-study scores of software-tools for all participants. Red is the score from CogState, green is the score from hamet e and purple is the score from GATRAS.

With score normalization, one can directly compare the results from one test with another even though the scoring methods for each test is different. With stacked column diagrams such as the diagram seen in **Figure 48**, each participant has a different color assigned to the score of the different tests such as red for CogState and green for hamet e. If the size of each color is similar to one another (i.e. each color takes up ~33% of the total length of the bar), it can be said that the participant performed consistently in each test. For example, an optimal result would be if a participant had the 4th highest score for the CogState test, he or she would also have around the 4th highest score for the other tests. However, if there is a significant size difference in the length of each colored, then they performed significantly better with one test than the other. An example can be seen in **Figure 48**, where Participant 4 performed significantly better in the GATRAS tests in comparison with the other test

As can be seen in **Figure 48**, the scores of the participants were quite constant. For most participants each color represented between 25% and 40% of their score with the vast majority with sizes between 31% and 36%. This shows a strong relationship between the three software-tools and suggests that if one does well with one of the software-tools, he or she would also do well with the others.

The tasks from each software were also separated into three categories in order to determine whether or not tests that assess similar cognitive functions such as memory or decision making from different software-tools correlate with one another.

Memory-based Tests

The first category that was evaluated was the memory-based tests category. As the name implies, all games in this category assess participant's ability to remember paths, shapes, positions and instructions. This category consists of the One Card Learning and Groton Maze Tasks from CogState, Write SMS and Weigh Package Tasks from hamet e and the Monkey Ladder Game from GATRAS.

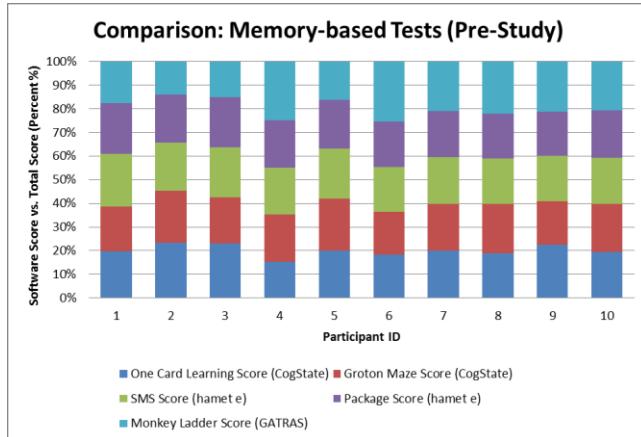


Figure 49. Normalized pre-study scores of all memory-based tests for all participants. Blue is the One Card Learning Task from CogState, red is Groton Maze Task from CogState, green is Write SMS from hamet e, purple is Weigh Package from hamet e and light blue is Monkey Ladder Game from GATRAS.

For the memory-based tests, an ideal distribution of each color on the bar is 20%. That is, with ideal results, each color should take up exactly 20% of the total bar length. From the results in **Figure 49**, one can see that the scores of the participants in memory-based tests in the pre-study are close to the ideal distribution. With the exception of participants 4 and 6, most of the participants managed to have a distribution of $20\% \pm 3\%$. This implies, for example, that if a user gets the top score in one of the memory-based tests, he or she would also get the top score in the other memory-based tests.

Decision-Making-based Tests

The tasks in the second category, decision-making-based tests, assess the participant's ability to make decisions such as choosing the right path or choosing the correct answer from a list of possible solutions. In this category the tasks are: Detection and Identification Tasks from CogState, Weigh Package Task from hamet e and Maze, Combination and Rotation Games from GATRAS.

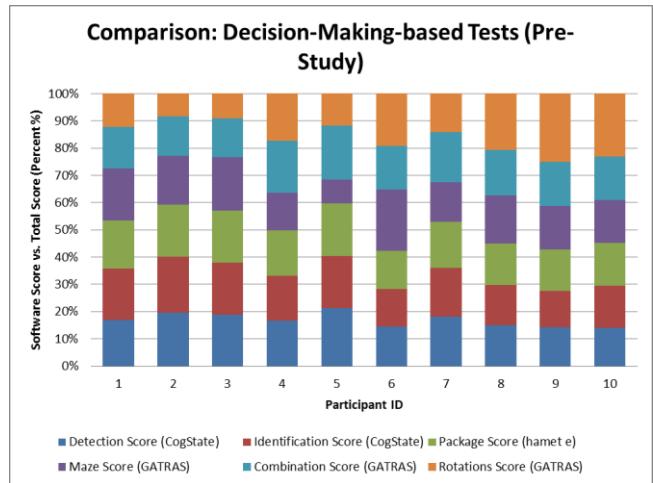


Figure 50. Normalized pre-study scores of all decision-making-based tests for all participants. Blue is the Detection Task from CogState, red is the Identification Task from CogState, green is Weigh Package from hamet e, purple is the Maze Game from GATRAS, light blue is the Combination Game from GATRAS and orange is the Rotation Game from GATRAS.

Unlike the results from the memory-based tests, the distribution in **Figure 50** does not show a good correlation between the tests in the decision-making-based category. As one can see in **Figure 50**, many participants performed extremely well in the Maze Game in GATRAS (purple) but then got a much lower rank in comparison to the other participants in the Rotation Game in GATRAS.

Learning-based Tests

The last category, learning-based tests, consists of tests which assess and evaluate the participant's ability to understand how to complete the task. An example of this the Data Transferral Task for hamet e where the participant needs to understand that they need to take the data from the physical list and input the missing information in the software without the tester having to tell the user what to input. This category consists of 8 tests: Groton Maze Task from CogState, Data Transferral, Write SMS, Input Coordinates and Weigh Package Tasks from hamet e and Combination, Shapes and Rotation Games from GATRAS.

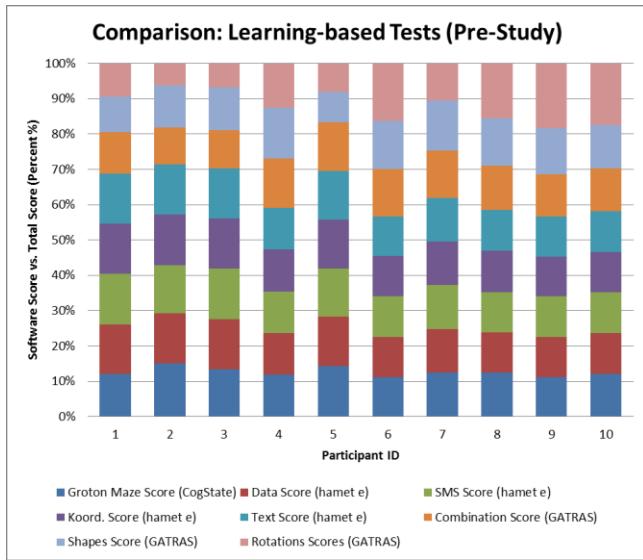


Figure 51. Normalized pre-study scores of all learning-based tests for all participants. Blue is the Groton Maze Task from CogState, red is Data Transferal from hamet e, green is Write SMS from hamet e, purple is Input Coordinates from hamet e, light blue is Type Words from hamet e, orange is the Combination Game from GATRAS, sky blue is the Shapes Game from GATRAS and pink is the Rotations Game from GATRAS.

With ideal results, the results of each learning-based test would take up 12.5% of the total bar. However, as seen in **Figure 51**, the results from the study fail to achieve this result. While the distribution of colors for each bar is less varied than those from the decision-making-based tests, there is still a variance of $\pm 7\%$ for the results in this category.

MAIN STUDY

As mentioned before, the main goal of the main study is to evaluate the features and usability of the software as well as see whether or not the results from the software-tools correlate with the results from the Ground Truth Study as well as the results from each other. The treatment group consists of participants with various disabilities from Gemeinnützige Werkstätten und Wohnstätten GmbH (GWW) in Sindelfingen, Germany. The participants will be asked to do a series of computerized assessment tests from GATRAS, developed by the University of Stuttgart; Cogstate, developed by CogState Research and the hamet e computerized tests developed by the Berufsbildungswerk Waiblingen. To assess the validity of the software test results, the workers' performance will be compared to the results from the Ground Truth Study, the results from each software-tool as well as the performance ratings gained from personal observations of the participants and questionnaire results filled out by the participants.

Procedure

Before the participants begin the tests, the participants were asked to fill out a questionnaire with questions ranging from documenting the participant's usual activity regimen to assessing the participant's current mental and physical state. Originally, the supervisors were asked to assist the participant in completing the survey due to fears of "acquiescence": the tendency to answer positively to questions whether or not they agree to the questions. However, due to time constraints and assurances that most of the participants would not have a problem filling out the questionnaire, the majority of the questionnaires were completed with the help of the test instructors instead of the participant's overseers.

After filling out the first three questions in the survey, the participants were then asked to do the tasks in either GATRAS, CogState or hamet e.

For GATRAS, before the participants began each test, the test instructors would explain to the participant what he or she was supposed to do in the test. Afterwards, the participants were allowed an unspecified amount of unrecorded training until the test instructors were confident that the participant understood the task or if the participant signaled that he or she understood the task and was ready for the recorded test. Due to the varying levels of cognitive disabilities between the participants, the time required for training normally ran between 30 and 90 seconds.

With CogState and hamet e, similar to GATRAS, the test instructors would first explain to the participant the goal of the test before the participant began. Afterwards, they were asked to complete the practice tasks included with the software. As these practice tasks were unskippable, the practice tasks could take between 60 seconds and 30 minutes to complete depending on the participant. After the practice task has been completed by the participant, the software would then automatically start the recorded trials after confirmation from the test supervisor.

While the participants are doing the test, the scores, time and mistakes for GATRAS and CogState and time and mistakes for hamet e are logged by the software. Furthermore, during this time, the test instructor would observe the participant and note reasons why a participant was doing particularly well or badly.

After the participant has completed all the tests in a particular software-tool, they were asked to fill out a section of the questionnaire (**Appendix Figure A2**) which attempts to understand how difficult the participant found the tests in the software-tool as well as whether or not they enjoyed doing the test. In addition, the questionnaire surveys the comfort and stress levels of the participant during the tests. This was necessary as being observed by strangers may influence the participant's motoric and cognitive abilities due to shyness or anxiety.

Subsequently, the participants were then asked to complete the next software-tools he or she had not yet completed as well as fill out the relevant questions in the questionnaire.

The scores, numbers of errors made as well as the results of the survey were then taken into account when calculating and evaluating the final results. These results were then compared to each other in addition to the results of the Ground Truth Study.

Differences between Pre- and Main Study

Due to the different cognitive and motoric abilities between the students at the University of Stuttgart and the participants from GWW, different procedures and materials were used in the implementation of both studies.

The main difference between the pre- and main study is the different time limit set for the GATRAS tests. The participants in the pre-study had a time limit of 2 minutes per game in GATRAS compared to the time limit of 3 minutes given to the participants of the main study. This was done due to several factors: expected scores, understanding and motivation. One of the main factors why the time limit was lowered for the students from the University of Stuttgart was because their expected scores per minute were predicted to be extremely high and their expected mistakes to be very low. These assumptions were made due to the fact that the majority of the pre-study participants were young and were majoring in technical fields such as computer science or engineering. This meant that the participants would be above average in terms of "tech-savviness", their proficiency with computers or related technologies, with nearly all the pre-study participants having had experience with touch-based devices through constant everyday use. All students that participated in the pre-study had also been studying at the university for several years. Therefore, it was theorized that the vast majority of the 10 pre-study participants would be able to understand each game well under the 30 second learning time limit apposed upon them and that they would find the games extremely simple. This reasoning also ties in with the third factor, motivation. It was hypothesized, that the participants of the pre-study would find the games extremely boring and repetitive if done longer than two minutes due to the simplicity of the games. If the game time was increased, the majority of the students would lose their interest and motivation as they get tired of doing repetitive actions and therefore indirectly influence their results.

Additionally, the participants in the main study had to first participate in the Ground Truth Study before they could take part in the main study whereas the pre-study participants were not required to do the Ground Truth Study before participating in the pre-study. However, this was mainly due to the time constraints of the pre-study participants as the majority were full time students or had part time jobs.

There were also some differences between the questionnaires given to the pre- and main study participants. While the majority of the questions remained the same, the two studies had different methods of receiving feedback from the participant. With the pre-study, the participants were asked for written and vocal feedback which allows the participant to describe in detail the reason why they thought they made mistakes or why they were uncomfortable in addition to provide constructive feedback over the entire process. This allowed us to view in detail what procedures did and did not work. With the students in the pre-study, the assumption could be made that all participants would be able to read and write as well as have a high level of creative thinking. However, with the participants in the main study, such an assumption could not be made as many of the participants could have severe cognitive disabilities. Therefore, the feedback questions in the main study questionnaire were fully multiple-choice with several pre-determined options listed as default. The participants in the main study were also encouraged, to mark the "other" option if they felt it was appropriate and explain verbally to one of the tester's or his or her supervisor his or her reasoning.

Evaluation

Here, the different software-tools (GATRAS; CogState and hamet e) are evaluated in order to determine whether or not the results from the software-tools correlated with the real-world Ground Truth Study as well as assess whether or not the participant liked the software-tools, determine the difficulty of each software as well as collect feedback for future studies.

Questionnaire Analysis

As with the pre-study, the questionnaire was divided into four parts: the first part was completed by the participants before the actual tests began. The second, third and fourth part of the questionnaire was filled out by the participant after he or she completed the GATRAS, CogState and hamet e tests respectively.

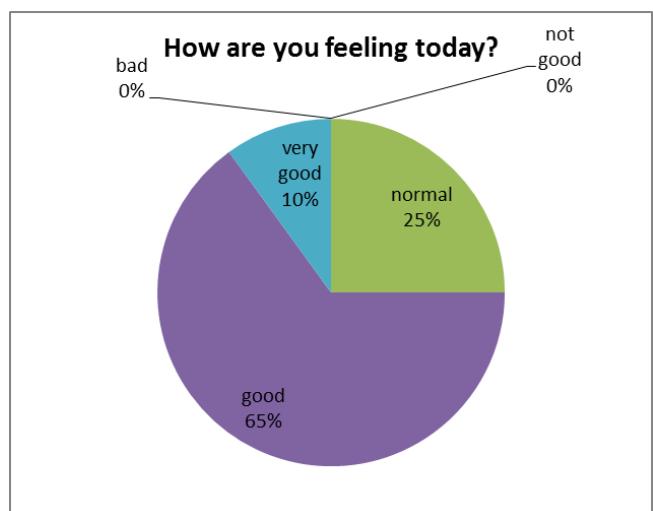


Figure 52. Results from Question 1 from the main study questionnaire.

As with the pre-study, the first question served as a way to measure the motivation and condition of the participants. As shown in **Figure 52**, none of the participants were feeling unwell and 75% or 15 of the 20 participants were feeling above average on the day of testing. As the vast majority of the participants were similarly motivated and in roughly the same condition compared to their average mood, it can be assumed that their motivation and condition affect their results in approximately the same way.

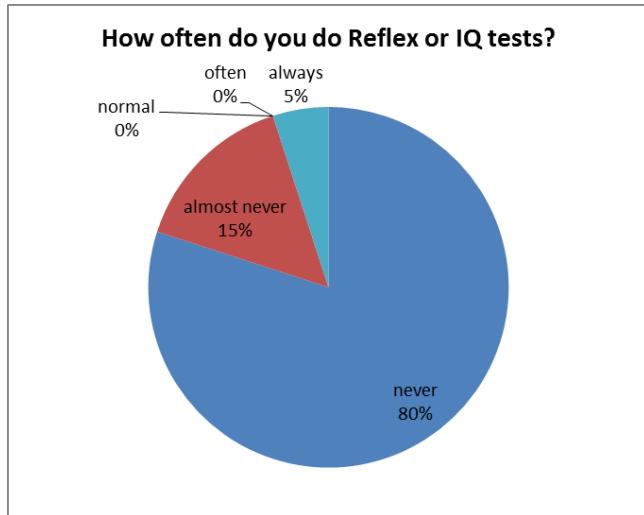


Figure 53. Results from Question 2 from the main study questionnaire.

The second question “How often do you do Reflex or IQ test?” show that 80% of the participants have never done such tests before. This could mean that the majority of participants have never seen or interacted with such tests before. This meant that a large proportion of the test population were interacting with such tests for the first time which could hypothetically increase the time needed to understand and learn the rules and tasks. This should especially be considered when reviewing the results from the main study as GATRAS and to a certain extent CogState tests are partially based on reflex and IQ tests whereas the hamet e tests are based on simulation tasks.

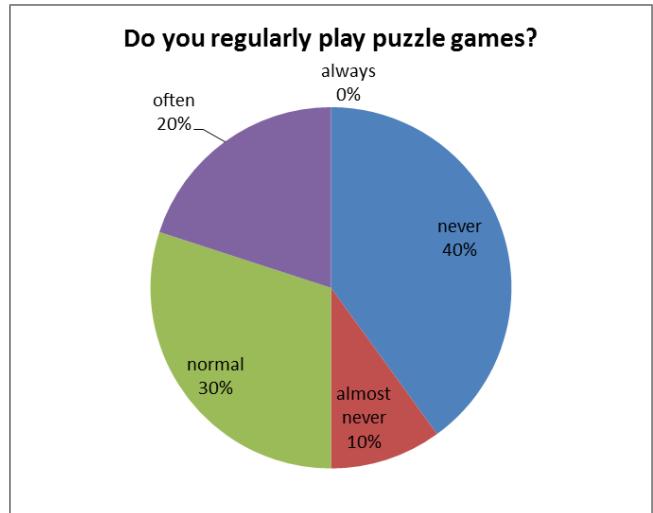


Figure 54. Results from Question 3 from the main study questionnaire.

As can be seen in **Figure 54**, 50% of the participants regularly play puzzle games. As puzzle games are primarily cognitive games, the high number of participants that have experience with puzzle games could potential distort the results in favor for the participants with puzzle game experience.

The questions in the second, third and fourth part of the questionnaire assesses the condition of the participant during the tests as well as how difficult they found each software and whether or not they would do such a test again.

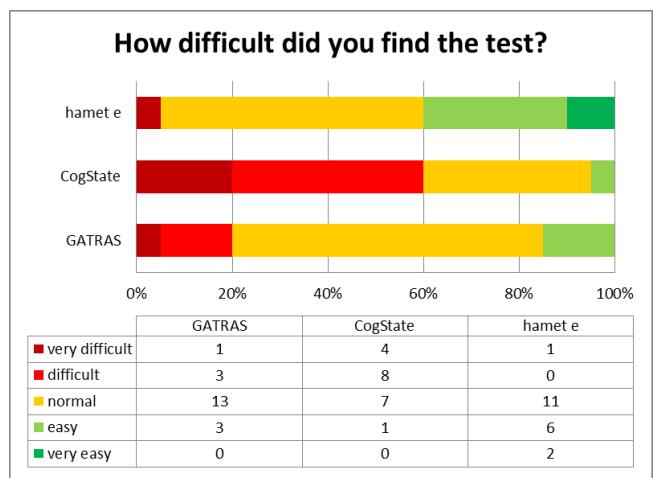


Figure 55. Results from the question “How difficult did you find the test?” for each of the software-tools.

As can be seen in **Figure 55**, the vast majority of the participants found the CogState tasks to be the most difficult. Out of 20 participants, 12 participants rated the CogState tests as difficult or very difficult whereas GATRAS had 4 and hamet e only 1 participant which rated the tests as difficult or very difficult. As further proof of the

difficulty of the CogState tasks, only 1 participant said that the CogState software was easy to complete versus the 8 from hamet e and 3 from GATRAS. However, the participant that listed the CogState software as easy also listed the other software-tools as easy. Therefore, this result could be due to the participant losing interest while completing the questionnaire or due to “acquiescence”, the tendency to answer positively to questions whether or not they agree to the questions.

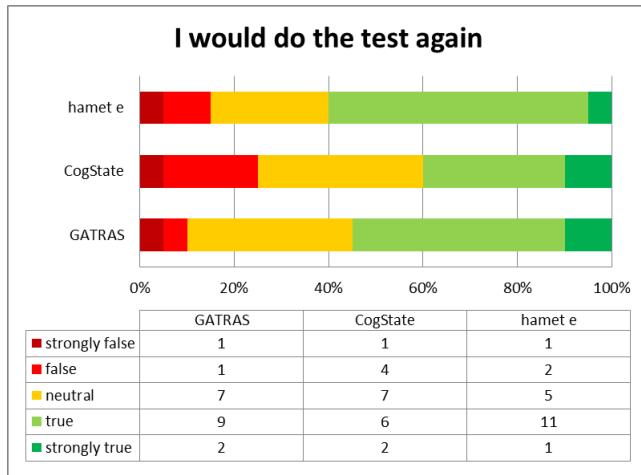


Figure 56. Results when asking the participants whether or not they would do the test again.

Similar to the question where the participant needed to rate the difficulty of each software, most of the participants decided that they would do the hamet e and GATRAS tests again whereas only 40% of the participants would do that CogState tests again.

Results Analysis

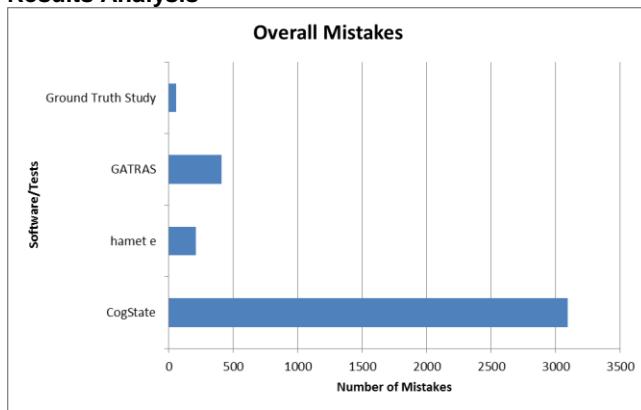


Figure 57. Total number of mistakes made by the participants for each software/test.

Software/Test	Mistakes
CogState	3094
hamet e	211
GATRAS	411
Ground Truth	57

Table 2. Total number of mistakes made by the participants for each software/test.

Similar to the pre-study results, **Figure X6** and **Table X1** shows that of the three softwares, the participants made the most mistakes in the CogState tests. In CogState, the participants made in total over 3000 mistakes versus the 679 mistakes made in the other three tests combined. That is 82% of all mistakes made in the study. This vast disparity in the number of mistakes between CogState and the other two software-tools indicates just how difficult the CogState test is in comparison to the softwares developed by the University of Stuttgart and the Berufsbildungswerk Waiblingen.

ID	CogState	C. Normalized	Hamet	H. Normalized
1	203	63.083	61.5	17.260
2	196	60.907	326.6	91.662
3	327	101.616	394.8	110.802
4	366	113.735	430	120.681
5	358	111.249	297.4	83.467
6	391	121.504	465.8	130.729
7	402	124.922	452.8	127.080
8	327	101.616	378.2	106.144
9	257	79.863	336.6	94.468
10	357	110.938	396.1	111.167
11	348	108.142	437.6	122.814
12	268	83.282	120.7	33.875
13	379	117.775	419.8	117.819
14	340	105.656	447.8	125.677
15	330	102.548	372.1	104.432
16	246	76.445	278.9	78.275
17	372	115.600	450.1	126.323
18	347	107.831	357.4	100.306
19	294	91.361	431.8	121.187
20	328	101.927	270.2	75.833

Table 3. Table of raw and normalized scores for CogState and hamet e.

ID	GATRAS	G. Normalized	Scheren Score	S. Normalized
1	72	24.628	8	11.241
2	371	126.903	47	65.903
3	272	93.039	74	103.843
4	302	103.301	85	119.441
5	261	89.277	90	126.467
6	430	147.084	97	136.443
7	463	158.372	82	114.523
8	300	102.617	56	78.971
9	237	81.067	73	102.579
10	388	132.718	86	120.424
11	337	115.273	88	124.218
12	128	43.783	34	48.338
13	374	127.929	78	109.042
14	426	145.716	71	99.206
15	238	81.409	91	128.434
16	162	55.413	81	113.398
17	502	171.712	67	94.709
18	162	55.413	74	103.843
19	315	107.748	58	81.782
20	107	36.600	83	117.192

Table 4. Table of raw and normalized scores for GATRAS and the Ground Truth Study (Scheren Score).

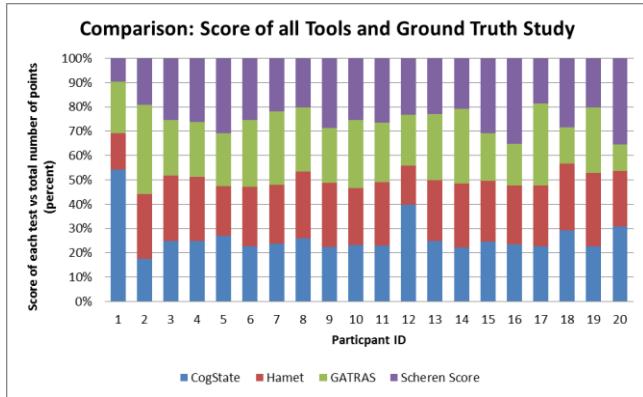


Figure 58. Normalized scores of software-tools for all participants. Blue is the score from CogState, red is the score from hamet e, green is the score from GATRAS and purple is the score from the Ground Truth Study.

As one can see in **Figure 58**, while many of the participants had “relatively” constant scores in each test, there were several extreme outliers such as Participants 1, 2 and 12 for CogState or 18 and 20 for GATRAS. However, the hamet e scores have stayed relatively constant in comparison with the scores from GATRAS and CogState as the size of each red bar is similar across nearly all participants. This means that many of the participants performed “as expected” for the hamet e tests.

$$\text{unity-based normalization} = \frac{\text{score}(\text{participant ID}) - \min(\text{score}(\text{all}))}{\max(\text{score}(\text{all})) - \min(\text{score}(\text{all}))}$$

Formula 3. Equation for computing unity-based normalized scores.

For the following scatter plot diagrams, **Formula 3** was used to put all the results in the same scale so that an objective comparison would be possible.

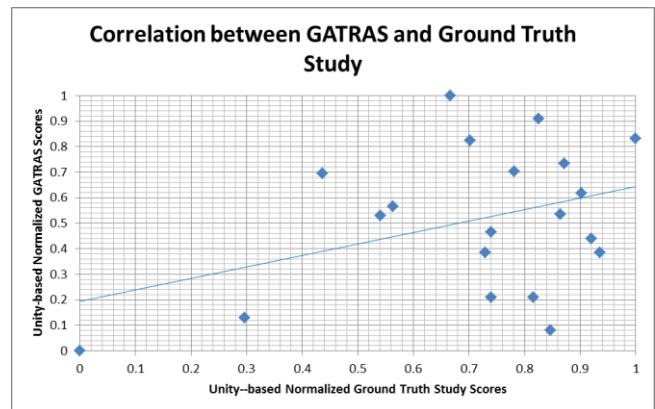


Figure 59. Scatter plot diagram of unity-based normalized GATRAS scores in contrast to unity-based normalized scores from the Ground Truth Study. $R^2 = 0.1455$.

As one can see in **Figure 59**, one can see a correlation between the Ground Truth Study Scores and those from GATRAS. This can be seen in the upward trend of the trend line meaning that on average, if one does better in the shear production test, then he or she should get a better score in GATRAS or vice versa.

The correlation strength between two sets of data is dependant on the correlation coefficient (R^2). The closer the correlation coefficient is to 1 for positive trend lines or -1 for negative trend lines, the stronger the correlation. However, **Figure 59** only has a correlation coefficient of 0.1455 which shows that the strength of the correlation between GATRAS and the Ground Truth Study is actually quite low. It should be noted that some of the participants had experience working in production lines which could explain some of the large variances between the GATRAS scores when compared to the shear production times.

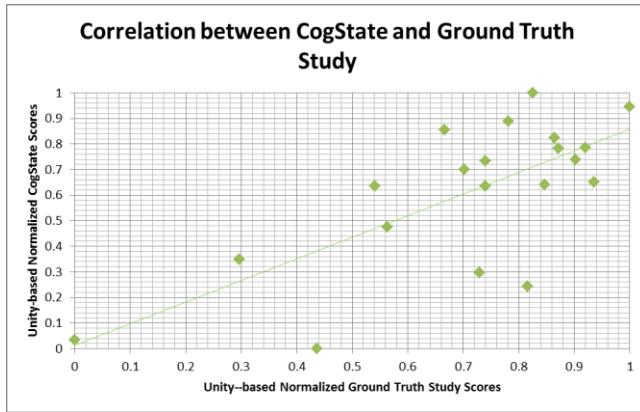


Figure 60. Scatter plot diagram of unity-based normalized CogState scores in contrast to unity-based normalized scores from the Ground Truth Study. $R^2 = 0.5068$

Similar to the scatter plot diagram between GATRAS and the shear production test, there is also a positive trend between the scores from CogState and those from the Ground Truth Study. In addition, as the variances between the CogState scores in **Figure 60** are much smaller than those of GATRAS in **Figure 59** and the data points are more clustered around the trend line, one can assume that the CogState scores have a better correlation with the Ground Study Truth than the GATRAS scores. This is further proven by the correlation coefficient between the CogState results and those from the Ground Truth Study. In comparison with the correlation coefficient from **Figure 59**, **Figure 60** has a much better correlation strength of 0.5068. While a correlation coefficient of 0.5068 means that the correlation between CogState and the Ground Truth Study is of middling strength, it is still over three times stronger than the correlation between GATRAS and the Ground Truth Study.

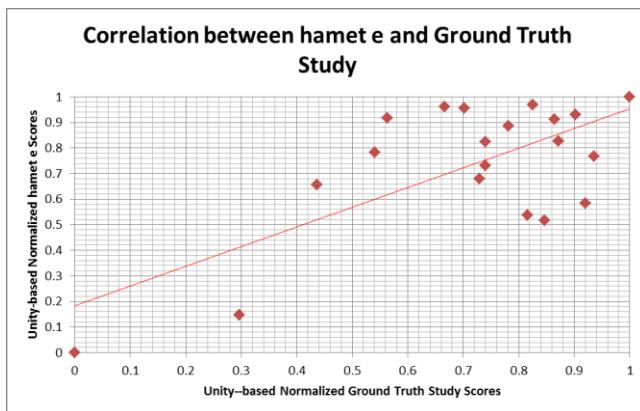


Figure 61. Scatter plot diagram of unity-based normalized hamet e scores in contrast to unity-based normalized scores from the Ground Truth Study. $R^2 = 0.4827$

As with GATRAS and CogState, a comparison of hamet e and Ground Truth Study scores (**Figure 61**) also shows a

positive upward trend in the trend line. However, the variance between the points is much smaller than those in the GATRAS and CogState comparisons. From these observations, the results suggest that of the three software, hamet e scores correlate the best with the results from the shear production. That is, if one does well in the Ground Truth Study, there is a high probability that he or she will do well with the hamet e software. This is further shown in the correlation strength of the two data sets (hamet e and Ground Truth Study) as correlation coefficient is 0.4827. While it is slightly lower than that of CogState and the Ground Truth Study, it still shows significant correlation strength in comparison with the GATRAS software-tool.

In general, all three software-tools show a positive trend line which meant that all three software-tools correlated with the Ground Truth Study. However, the GATRAS software had poor correlation strength whereas the hamet e and CogState software had moderate correlation strength. Nevertheless, from the results, one can surmise that the better a participant does in hamet e, CogState and GATRAS, the better his or her results from the shear production test. However, this effect is less visible with GATRAS.

Memory-based Tests

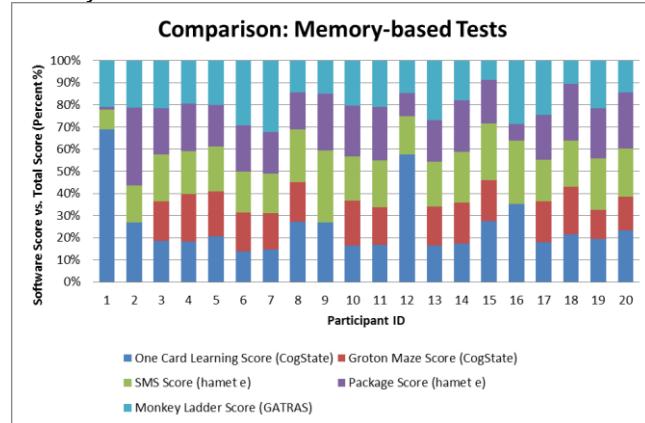


Figure 62. Normalized scores of all memory-based tests for all participants. Blue is the One Card Learning Task from CogState, red is Groton Maze Task from CogState, green is Write SMS from hamet e, purple is Weigh Package from hamet e and light blue is Monkey Ladder Game from GATRAS.

As one can see from **Figure 62**, the red bar for the Groton Maze Task is missing for some of the participants. This is either because they decided to drop out because of the difficulty or because the test timed out because the participant took too long to complete the task. However, for the participants that did take part in the Groton Maze Task, the size of the bars were quite constant meaning that they performed at around their theoretical norm in this task. On the other hand, one can see a huge disparity in scores between the One Card Learning Task (blue bar) and the

other games. This could be because it is easier to get a good score in the One Card Learning Task as you have a 50% chance to get the correct answer if one does not remember if they have seen this card before or not.

Decision-Making-based Tests

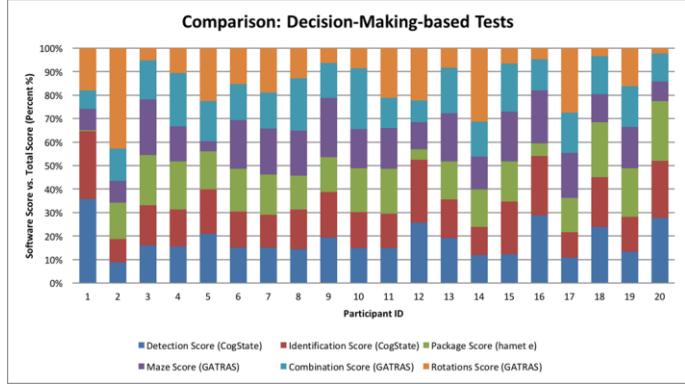


Figure 63. Normalized scores of all decision-making-based tests for all participants. Blue is the Detection Task from CogState, red is the Identification Task from CogState, green is Weigh Package from hamet e, purple is the Maze Game from GATRAS, light blue is the Combination Game from GATRAS and orange is the Rotation Game from GATRAS.

Unlike with memory-based tests, one can clearly see from **Figure 63**, a huge disparity between each test scores (especially in the Detection Task and Rotation game). As can be seen in the graph, many of the participants either performed extremely well in the Detection Task by CogState or the Rotation Game by GATRAS in comparison with their other tests or extremely poorly in comparison with their other tests. Therefore from this graph alone we can already see that the correlation between each test in this category is quite low as each participant performed unequally in each test. This can be seen later on when we perform the t-test in order to attempt to prove correlation between the tests in the category.

Learning-based Tests

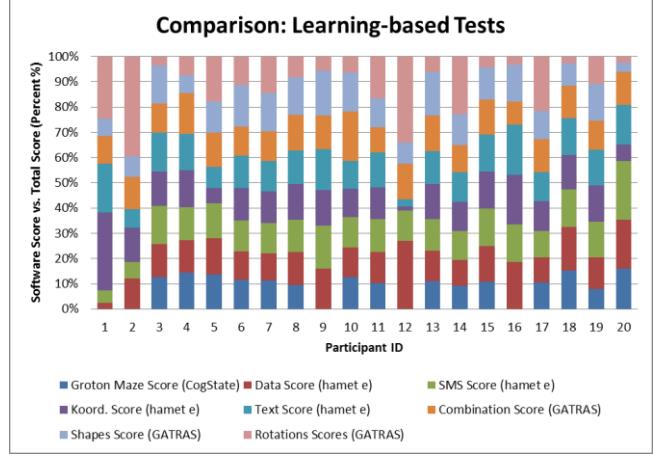


Figure 64. Normalized scores of all learning-based tests for all participants. Blue is the Groton Maze Task from CogState, red is Data Transferal from hamet e, green is Write SMS from hamet e, purple is Input Coordinates from hamet e, light blue is Type Words from hamet e, orange is the Combination Game from GATRAS, sky blue is the Shapes Game from GATRAS and pink is the Rotations Game from GATRAS.

As can be seen in **Figure 64**, while some participants performed relatively constantly compared with the other participants across all tests in this category such as participants 6, 7 and 19, many of results are similar to the decision-making-based tests in that the vast majority of the participants did not perform similarly for each of the tests in the category. The theoretical norm for each test is 12.5% which as you can see on the graph is something the results fail to achieve.

Data Analysis with t-Tests

While the correlation between individual participants and games are important, it is also important to evaluate the relevance between the tests. One common approach in evaluating correlations of data sets is to calculate a student's t-test. The t-Tests used to analyze the scores are the t-tests for evaluating paired two samples.

For the t-test, we assume the following two hypotheses:

Null hypothesis (H0): There are no similarities between the data sets, i.e. the participants did not perform equally between the two data sets.

Alternative hypothesis (HA): There are similarities in the two data sets, i.e. the participants performed equally in both tests.

Furthermore, we assume an alpha value of 0.1. That means if both the one-tail and two-tail values from the t-test are less than the alpha value, then the null hypothesis is accepted and the alternative hypothesis is not accepted since the probability that the data sets correlate due to random chance is greater than 90%.

All Software-tool Tests

In this section, the results from each software-tool is compared with one another and with the results from the shear production test through two sample t-Tests in order to find out if there is a correlation between said data sets. All t-Tests can be found in the **Appendix**.

As can be seen in **Tables A1** and **A2**, the scores from CogState correlates well with the results from hamet e, GATRAS vice versa. This is due to the one-tailed and two-tailed p-values for all three t-Tests being over the set alpha value of 0.1. This implies that for all three t-Tests, the null hypothesis has been rejected. The participants performed similarly enough in the software tests to say that the correlation is not induced by chance.

Unfortunately, the results from the rest of the t-Tests (**Table A3** to **Table A6**) suggests that the rest of the data sets do not correlate with one another as both the one-tailed and two-tailed p-values for the t-Tests are all less than the set alpha value of 10%. This implies that the null hypothesis has then been accepted as there is not enough data to suggest that the participants performed equally in each of the assessed tests.

Memory-based Tests

The t-Tests in this section attempts to find a correlation between each of the software-tools' memory-based tests as well as a correlation between the memory-based tests and the results from the Ground Truth Study.

As seen in **Table A7**, the $P(T \leq t)$ one-tail and two-tail is less than 0.1. This means that the null hypothesis is accepted as the probability that the results correlate due to random probability is greater than 90%. Thus the scores from the tests in CogState and hamet e which evaluates the memory function of the participants do not correlate with one another.

This procedure has also been applied to CogState vs. GATRAS and hamet e vs. GATRAS with the following results:

Similar to the results from the t-Test between CogState and hamet e, both the $P(T \leq t)$ one-tail and two-tail is less than 0.1 for both **Table A8** and **Table A9**. Therefore, it is unlikely that there is a correlation between the memory-based tests from CogState and GATRAS and between hamet e and GATRAS.

As can be seen in **Table A10** and **Table A11**, both the one-tail and two tail p-values for both these tables are greater than 0.1. This suggests that there is a relatively strong correlation between the results from the memory-based tests from CogState and hamet e and the results from the Ground Truth Study. This implies that if the user does well in one of these software-tools, then the user would also do well when manufacturing shears and vice versa. In contrast, the results from the memory-based tests for GATRAS do

not correlate well with the results from the Ground Truth Study as shown in **Table A12**.

Decision-Making-based Tests

The t-Tests in this section attempts to find a correlation between each of the software-tools' decision-making-based tests as well as a correlation between the decision-making-based tests and the results from the Ground Truth Study.

Unlike with memory-based tests, the t-Test between the decision-making tests from CogState and hamet e (**Table A13**) show a correlation as both the one-tailed and two-tailed p-values are over 0.1. This means that any correlation in the results being due to random chance is unlikely.

Similar to the memory-based tests between CogState and GATRAS, the results from the decision-making-based tests between those CogState and GATRAS (**Table A14**) also show little or no correlation as both the one and two-tailed p-values are under 0.1.

On the other hand, as can be seen in **Table A15**, the t-Test between the decision-making-based tests from hamet e and GATRAS shows a low correlation between the two software-tools as the one-tailed p-value is under 0.1 but the two-tailed p-value is over 0.1.

In contrast to the results of the memory-based tests between CogState and the Ground Truth Study, the t-Test between the decision-making-based tests from CogState and Ground Truth Study (**Table A16**) shows little to no correlation between the two data sets as both p-values are under 0.1.

On the other hand, both hamet e and GATRAS in particular, correlated well with the results from the shear production test (**Tables A17** and **A18** respectively). The one-tailed and two-tailed p-values between hamet e and the Ground Truth Study was slightly over the alpha value of 0.1 whereas the p-values for the t-Test between GATRAS and the Ground Truth Study were well over the 0.1 limit ($P(T \leq t)$ one tail 0.353 and $P(T \leq t)$ two tail 0.707).

Learning-based Tests

The t-Tests in this section attempts to find a correlation between each of the software-tools' learning-based tests as well as a correlation between the learning-based tests and the results from the Ground Truth Study.

Unlike with the t-Test results from the memory-based tests, all the learning-based tests seem to correlate with one another as both the one-tailed and two-tailed p-values for all three software-tool comparisons are well over 0.1 (**Tables A19** to **A21**). This suggests that the observed data is consistent with the assumption that our null hypothesis, that there is no correlation between the tests, is false.

For the t-Tests between the results of the three learning-based tests and those from the Ground Truth Study, all three software-tools correlated well with the Ground Truth Study as all three t-Tests (**Table A22** to **A24**) had p-values

well over 0.1. This means that for all three t-Test reject the null hypothesis that there are no similarities between the software-tools and the Ground Truth Study and accepts the alternative hypothesis that the participants performed equally in the software tests and in the shear production test. Therefore, these t-Test results suggest that if someone does well in tests that assess the participant's ability to learn and understand, he or she will also do well when producing shears and vice versa.

CONCLUSION

During the course of the study, we try to answer the question "which software is the most suitable to assess and evaluate the cognitive ability of an impaired person". To do this, we selected three software-tools that evaluated the various areas of cognitive function such as memory function or decision-making ability of people with neurological impairments. The tests and features of each software were analyzed and compared with one another. In addition, two studies were performed, one with students from the University of Stuttgart and another with the workers of the Gemeinnützige Werkstätten und Wohnstätten GmbH in order to receive feedback on the usability of the software-tools from the view of the participants as well as compare the results of the software-tools with one another to see if there is a correlation between each software. That is, to see if the participants would perform equally in every software instead of doing much better in one when compared with the other participants. However, in the end, we decided that none of the three software-tools could be recommended to potential users at their stages. This is due to several factors such as the various drawbacks and benefits of all three software and that there is no software that is decisively better than the others in general.

Because GATRAS is still in its development stage, it is still possible to add features and fine-tune the program which is a definite advantage for the software-tool. Furthermore, because of its level of gamification, the participants found the games enjoyable and could therefore enjoy a stable level of motivation throughout the test. Furthermore, the GATRAS tests are extremely plannable as they have a programmable time limit that does not require the participant to complete every single level of every test. However, a general disadvantage of the GATRAS software is that it does not have fully-fledged data analysis tools. While none of the three software-tools have the ability to save and analyze the test data in Excel or an Excel-alternative format, the score analyzer from GATRAS only shows that score and number of mistakes for every game in TXT-format. This is a problem as it is very difficult to read and all the statistics such as accuracy, reliability and result comparisons have to be manually calculated by the user.

If the paper just analyzed and compared the features of these three software-tools that assessed the ability of people with impairments, CogState would definitely be the

recommended product. The software itself is very professionally made. The user interface is easy to read and procedures for each test are pre-written and includes instructions for the test instructors as well as the participants. Furthermore, it has a large variety of data analysis tools and options such as improvement and action tracking which is very useful for people hoping to analyze the data from a study. In addition, it has functions that allow test data from multiple computers to be easily merged to one online cloud location with just a single click. However, the main reason why we could not recommend this software is due to the difficulty of the tasks that the participants needed to complete. The Groton Maze Learning Task was so difficult that 6 participants either timed-out from taking too long to complete the task or decided he or she did not want to continue with this task and asked to move on to the next task instead. Furthermore, the One Card Learning Task was too long for the participants and many just randomly clicked or just clicked "yes" or "no" to finish the test faster after the first 5 minutes.

Last of all, hamet e also has many advantages over the other two software-tools. All tasks are based on everyday computer office tasks that may come up in a company office. This allows the participants to train for work and the test instructors to see how well the participants will do in particular office tasks while assessing their neurological abilities at the same time. Furthermore, hamet e has different levels for every task which allows a wider variety of participants with cognitive impairments. However, the hamet e tasks are not suitable for people who do not have experience using computers or participants who cannot read as many of the tasks require the participant to read in order to proceed. Furthermore, the lack of a scoring system makes it too difficult for the users to analyze the test data as there is no clear value to help the analyzer rank and compare the participants as well as compare the results to a standardized baseline.

T-TEST COMPARISON		CATEGORIES			
Test 1	Tests 2	All Tests	Memory-based Tests	Decision-Making-based Tests	Learning-based Tests
CogState	hamet e	YES	NO	YES	YES
CogState	GATRAS	YES	NO	NO	YES
hamet e	GATRAS	NO	NO	YES (but low)	YES
CogState	GTS	NO	YES	NO	YES
hamet e	GTS	NO	YES	YES	YES
GATRAS	GTS	NO	NO	YES	YES

Table 5. Summary of t-Test results. “YES” means that the results from the two tests correlated with one another. “NO” means that they did not.

As for the correlation between the test data as well as those from the Ground Truth Study, we have received mixed results. While the results for the comparison between learning-based tests were very positive as every single software-tool correlated well with one another as well as with the Ground Truth Study, at least half of the t-Tests failed for each of the other categories. The results from the t-Tests indicated that only CogState correlated with hamet e and GATRAS (and vice versa) when taking all tests into account whereas the rest of the results from the “All Tests” category did not significantly correlate with each other. Furthermore, with memory-based tests, only those from CogState and hamet e correlated with the results from the Ground Truth Study. In the decision-making-based tests category, the scores from all tests correlated with one another except for the results from CogState and GATRAS and CogState and Ground Truth Study. These results suggest that if one wanted to use a real world test such as the Ground Truth Study as a baseline for tests, one would carry out learning-based tests as all the participants performed similarly or as expected in relation to each of the tests in this category.

To conclude, each of the three software-tools has features and advantages the other two software-tools lack as well as disadvantages that simply cannot be ignored. While the results are not definite as they still need to be confirmed through future studies with varying environments and larger control and treatment groups, from the results obtained in this study, an ideal software should encompass learning-based tests as the core of its cognitive assessment function as the study results show the highest correlation between the results in this category. Additionally, the tests should be highly gamified such as the tests from GATRAS and hamet e as they show the highest satisfaction and enjoyment for the participants. Furthermore, features such as the ability to save test data in an Excel or Excel-alternative format should have a high priority in future versions of cognitive assessment software as this will greatly simplify and reduce the time needed to analyze the test data from future studies. In addition, data analysis and organizational tools such as those provided by the CogState software should be implemented in such a software as well.

ACKNOWLEDGEMENTS

We would like to offer special thanks to Dr. Oliver Korn at the University of Stuttgart for his assistance, advice and encouragement during the preparation and implementation of the project. We would also like to extend our thanks to the staff of Gemeinnützige Werkstätten und Wohnstätten GmbH for allowing us to conduct our study at the company as well as organizing the locations and participants. Last but not least, we would like to express our great appreciation for the participants who took part in our studies for making it such an enjoyable experience.

APPENDIX

motionEAP: Hauptstudie Fragebogen		Alter:	Geschlecht (m/w):						
Software-Tools für Bewertung Kognitiven Fähigkeiten		Abteilung:			Nr.:				
1	Vor dem Experiment								
1	Wie geht es Ihnen heute?	schlecht	nicht gut	normal	gut				
		gar nicht	fast gar nicht	normal	oft				
2	Machen Sie Reflex- oder IQ-Tests oft?								
3	Spielen Sie Denkspiele (Puzzle, Rätsel, etc.) auf regulärer Basis?								
2	Nach GATRAS								
4	Haben Sie während des Tests Fehler gemacht? Falls ja, aus welchem Grund glauben Sie kam es zu dem Fehler?								
5	Würdest du in den Spielen etwas anderen? Wenn ja, was?								
6	Ich würde die Spiele wieder spielen.								
7	Ich würde die Spiele Kollegen empfehlen.								
8	Um die Spiele zu benutzen, muss mir jemand helfen.								
3	Nach CogState								
9	Haben Sie während des Tests Fehler gemacht? Falls ja, aus welchem Grund glauben Sie kam es zu dem Fehler?								
10	Würdest du in den Spielen etwas anderen? Wenn ja, was?								
11	Ich würde die Spiele wieder spielen.								
12	Ich würde die Spiele Kollegen empfehlen.								
13	Um die Spiele zu benutzen, muss mir jemand helfen.								
4	Nach Hamet								
14	Haben Sie während des Tests Fehler gemacht? Falls ja, aus welchem Grund glauben Sie kam es zu dem Fehler?								
15	Würdest du in den Spielen etwas anderen? Wenn ja, was?								
16	Ich würde die Spiele wieder spielen.								
17	Ich würde die Spiele Kollegen empfehlen.								
18	Um die Spiele zu benutzen, muss mir jemand helfen.								

Figure A1.

Pre-Study Questionnaire.

Questions 1-3 were filled out before the study began.

Questions 4-8 were filled out after the GATRAS tests have been completed.

Questions 9-13 were filled out after the completion of the CogState "Early Phase Battery".

Questions 14-18 were filled out after the Level 2 hamet e computerized tests have been completed.

motionEAP: Hauptstudie Fragebogen		Alter:	Geschlecht (m/w):		
Software-Tools für Bewertung Kognitiven Fähigkeiten		Abteilung:			Nr.:
1	Vor dem Experiment				
		schlecht	nicht gut	normal	gut
1	Wie geht es Ihnen heute?				
		gar nicht	fast gar nicht	normal	oft
2	Machen Sie Reflex- oder IQ-Tests oft?				
3	Spielen Sie Denkspiele (Puzzle, Rätsel, etc.) auf regulärer Basis?				
2	Nach GATRAS				
		sehr unbehaglich	unbehaglich	normal	behaglich
4	Haben Sie sich zu irgendeinem Zeitpunkt des Tests (Spiele, Oberfläche, Umgebung, Fragebogen, etc.) unbefriedigt gefühlt?				
		Umgebung	Schwerigkeit	Probleme mit dem Gerät	Probleme mit den Anweisungen
5	Haben Sie während des Tests Fehler gemacht? Falls ja, aus welchem Grund glauben Sie kam es zu dem Fehler?				
		sehr schwer	schwer	normal	einfach
6	Wie schwer fanden Sie die Spiele?				
		stimmt genau	stimmt	neutral	stimmt nicht
7	Ich würde die Spiele wieder spielen.				
8	Ich würde die Spiele Kollegen empfehlen.				
9	Um die Spiele zu benutzen, muss mir jemand helfen.				
3	Nach CogState				
		sehr unbehaglich	unbehaglich	normal	behaglich
10	Haben Sie sich zu irgendeinem Zeitpunkt des Tests (Spiele, Oberfläche, Umgebung, Fragebogen, etc.) unbefriedigt gefühlt?				
		Umgebung	Schwerigkeit	Probleme mit dem Gerät	Probleme mit den Anweisungen
11	Haben Sie während des Tests Fehler gemacht? Falls ja, aus welchem Grund glauben Sie kam es zu dem Fehler?				
		sehr schwer	schwer	normal	einfach
12	Wie schwer fanden Sie die Spiele?				
		stimmt genau	stimmt	neutral	stimmt nicht
13	Ich würde die Spiele wieder spielen.				
14	Ich würde die Spiele Kollegen empfehlen.				
15	Um die Spiele zu benutzen, muss mir jemand helfen.				
4	Nach Hamet				
		sehr unbehaglich	unbehaglich	normal	behaglich
16	Haben Sie sich zu irgendeinem Zeitpunkt des Tests (Spiele, Oberfläche, Umgebung, Fragebogen, etc.) unbefriedigt gefühlt?				
		Umgebung	Schwerigkeit	Probleme mit dem Gerät	Probleme mit den Anweisungen
17	Haben Sie während des Tests Fehler gemacht? Falls ja, aus welchem Grund glauben Sie kam es zu dem Fehler?				
		sehr schwer	schwer	normal	einfach
18	Wie schwer fanden Sie die Spiele?				
		stimmt genau	stimmt	neutral	stimmt nicht
19	Ich würde die Spiele wieder spielen.				
20	Ich würde die Spiele Kollegen empfehlen.				
21	Um die Spiele zu benutzen, muss mir jemand helfen.				

Figure A2.

Main Study Questionnaire.

Questions 1-3 were filled out before the study began.

Questions 4-9 were filled out after the GATRAS tests have been completed.

Questions 10-15 were filled out after the completion of the CogState "Early Phase Battery".

Questions 16-21 were filled out after the Level 2 hamet e computerized tests have been completed.

Two-Sample t-Test Assuming Unequal Variances		
	CogState	hamet e
Average	321.8	356.31
Variance	3519.95789	11821.1357
Observations	20	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	29	
t-Statistic	-1.24603946	
P(T<=t) one-tail	0.11136124	
t Critical one-tail	1.31143365	
P(T<=t) two-tail	0.22272248	
t Critical two-tail	1.69912703	

Table A1. Results of t-Test between the CogState and hamet e scores.

Two-Sample t-Test Assuming Unequal Variances		
	CogState	GATRAS
Average	321.8	292.35
Variance	3519.95789	15045.2921
Observations	20	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	27	
t-Statistic	0.96660691	
P(T<=t) one-tail	0.17115978	
t Critical one-tail	1.31370291	
P(T<=t) two-tail	0.34231957	
t Critical two-tail	1.70328845	

Table A2. Results of t-Test between the CogState and GATRAS scores.

Two-Sample t-Test Assuming Unequal Variances		
	CogState	GTS
Average	321.8	71.175
Variance	3519.95789	466.068289
Observations	20	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	24	
t-Statistic	17.7529003	
P(T<=t) one-tail	1.3139E-15	
t Critical one-tail	1.31783593	
P(T<=t) two-tail	2.6277E-15	
t Critical two-tail	1.71088208	

Table A3. Results of t-Test between the CogState scores and the results from the Ground Truth Study.

Two-Sample t-Test Assuming Unequal Variances		
	hamet e	GATRAS
Average	356.31	292.35
Variance	11821.1357	15045.2921
Observations	20	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	37	
t-Statistic	1.74509265	
P(T<=t) one-tail	0.0446351	
t Critical one-tail	1.30485438	
P(T<=t) two-tail	0.08927021	
t Critical two-tail	1.68709362	

Table A4. Results of t-Test between the hamet e and GATRAS scores.

Two-Sample t-Test Assuming Unequal Variances		
	hamet e	GTS
Average	356.31	71.175
Variance	11821.1357	466.068289
Observations	20	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	20	
t-Statistic	11.5037383	
P(T<=t) one-tail	1.4285E-10	
t Critical one-tail	1.32534071	
P(T<=t) two-tail	2.8569E-10	
t Critical two-tail	1.72471824	

Table A5. Results of t-Test between the hamet e scores and the results from the Ground Truth Study.

Two-Sample t-Test Assuming Unequal Variances		
	GATRAS	GTS
Average	292.35	71.175
Variance	15045.2921	466.068289
Observations	20	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	20	
t-Statistic	7.94193063	
P(T<=t) one-tail	6.5214E-08	
t Critical one-tail	1.32534071	
P(T<=t) two-tail	1.3043E-07	
t Critical two-tail	1.72471824	

Table A6. Results of t-Test between the GATRAS scores and the results from the Ground Truth Study.

Two-Sample t-Test Assuming Unequal Variances		
	CogState	hamet e
Average	76.025	62.035
Variance	1008.99936	1111.87721
Observations	40	40
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	78	
t-Statistic	1.92127713	
P(T<=t) one-tail	0.02917511	
t Critical one-tail	1.2924996	
P(T<=t) two-tail	0.05835021	
t Critical two-tail	1.66462464	

Table A7. Results of t-Test between the memory-based tests from CogState and hamet e.

Two-Sample t-Test Assuming Unequal Variances		
	CogState	GTS
Average	76.025	71.175
Variance	1008.99936	466.068289
Observations	40	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	52	
t-Statistic	0.69621561	
P(T<=t) one-tail	0.24469709	
t Critical one-tail	1.29804502	
P(T<=t) two-tail	0.48939418	
t Critical two-tail	1.67468915	

Table A10. Results of t-Test between the memory-based tests from CogState and the results from the Ground Truth Study.

Two-Sample t-Test Assuming Unequal Variances		
	Cogstate	GATRAS
Average	76.025	41.6
Variance	1008.99936	580.463158
Observations	40	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	49	
t-Statistic	4.67392259	
P(T<=t) one-tail	1.1717E-05	
t Critical one-tail	1.67655089	
P(T<=t) two-tail	2.3433E-05	
t Critical two-tail	2.00957524	

Table A8. Results of t-Test between the memory-based tests from CogState and GATRAS.

Two-Sample t-Test Assuming Unequal Variances		
	hamet e	GTS
Average	62.035	71.175
Variance	1111.87721	466.068289
Observations	40	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	54	
t-Statistic	-1.27859877	
P(T<=t) one-tail	0.10325321	
t Critical one-tail	1.29742649	
P(T<=t) two-tail	0.20650641	
t Critical two-tail	1.67356491	

Table A11. Results of t-Test between the memory-based tests from hamet e and the results from the Ground Truth Study.

Two-Sample t-Test Assuming Unequal Variances		
	hamet e	GATRAS
Average	62.035	41.6
Variance	1111.87721	580.463158
Observations	40	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	50	
t-Statistic	2.71096363	
P(T<=t) one-tail	0.00458487	
t Critical one-tail	1.67590503	
P(T<=t) two-tail	0.00916975	
t Critical two-tail	2.00855911	

Table A9. Results of t-Test between the memory-based tests from hamet e and GATRAS.

Two-Sample t-Test Assuming Unequal Variances		
	GATRAS	GTS
Average	41.6	71.175
Variance	580.463158	466.068289
Observations	20	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	38	
t-Statistic	-4.08849631	
P(T<=t) one-tail	0.00010849	
t Critical one-tail	1.3042302	
P(T<=t) two-tail	0.00021698	
t Critical two-tail	1.68595446	

Table A12. Results of t-Test between the memory-based tests from GATRAS and the results from the Ground Truth Study.

Two-Sample t-Test Assuming Unequal Variances		
	CogState	hamet e
Average	84.875	82.5
Variance	190.778846	1206.47368
Observations	40	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	22	
t-Statistic	0.29437163	
P(T<=t) one-tail	0.38561679	
t Critical one-tail	1.71714437	
P(T<=t) two-tail	0.77123359	
t Critical two-tail	2.07387307	

Table A13. Results of t-Test between the decision-making-based tests from CogState and hamet e.

Two-Sample t-Test Assuming Unequal Variances		
	CogState	GATRAS
Average	84.875	68.2333333
Variance	190.778846	2238.85989
Observations	40	60
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	73	
t-Statistic	2.56530691	
P(T<=t) one-tail	0.00618105	
t Critical one-tail	1.66599622	
P(T<=t) two-tail	0.0123621	
t Critical two-tail	1.99299713	

Table A14. Results of t-Test between the decision-making-based tests from CogState and GATRAS.

Two-Sample t-Test Assuming Unequal Variances		
	hamet e	GATRAS
Average	82.5	68.2333333
Variance	1206.47368	2238.85989
Observations	20	60
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	44	
t-Statistic	1.44381996	
P(T<=t) one-tail	0.07793632	
t Critical one-tail	1.68022998	
P(T<=t) two-tail	0.15587264	
t Critical two-tail	2.01536757	

Table A15. Results of t-Test between the decision-making-based tests from hamet e and GATRAS.

Two-Sample t-Test Assuming Unequal Variances		
	CogState	GTS
Average	84.875	71.175
Variance	190.778846	466.068289
Observations	40	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	27	
t-Statistic	2.58569347	
P(T<=t) one-tail	0.00771833	
t Critical one-tail	1.31370291	
P(T<=t) two-tail	0.01543666	
t Critical two-tail	1.70328845	

Table A16. Results of t-Test between the decision-making-based tests from CogState and the results from the Ground Truth Study.

Two-Sample t-Test Assuming Unequal Variances		
	hamet e	GTS
Average	82.5	71.175
Variance	1206.47368	466.068289
Observations	20	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	32	
t-Statistic	1.2384107	
P(T<=t) one-tail	0.11228608	
t Critical one-tail	1.30857279	
P(T<=t) two-tail	0.22457215	
t Critical two-tail	1.69388875	

Table A17. Results of t-Test between the decision-making-based tests from hamet e and the results from the Ground Truth Study.

Two-Sample t-Test Assuming Unequal Variances		
	GATRAS	GTS
Average	68.2333333	71.175
Variance	2238.85989	466.068289
Observations	60	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	70	
t-Statistic	-0.3778275	
P(T<=t) one-tail	0.353351	
t Critical one-tail	1.2937629	
P(T<=t) two-tail	0.706702	
t Critical two-tail	1.66691448	

Table A18. Results of t-Test between the decision-making-based tests from GATRAS and the results from the Ground Truth Study.

Two-Sample t-Test Assuming Unequal Variances		
	CogState	hamet e
Average	63.5	68.4525
Variance	1618.15789	1511.45417
Observations	20	80
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	29	
t-Statistic	-0.49574333	
P(T<=t) one-tail	0.31190585	
t Critical one-tail	1.69912703	
P(T<=t) two-tail	0.6238117	
t Critical two-tail	2.04522964	

Table A19. Results of t-Test between the learning-based tests from CogState and hamet e.

Two-Sample t-Test Assuming Unequal Variances		
	CogState	GATRAS
Average	63.5	69.1333333
Variance	1618.15789	2233.94802
Observations	20	60
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	38	
t-Statistic	-0.51828221	
P(T<=t) one-tail	0.30363312	
t Critical one-tail	1.68595446	
P(T<=t) two-tail	0.60726625	
t Critical two-tail	2.02439416	

Table A20. Results of t-Test between the learning-based tests from CogState and GATRAS.

Two-Sample t-Test Assuming Unequal Variances		
	hamet e	GATRAS
Average	68.4525	69.1333333
Variance	1511.45417	2233.94802
Observations	80	60
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	112	
t-Statistic	-0.09087829	
P(T<=t) one-tail	0.46387582	
t Critical one-tail	1.65857263	
P(T<=t) two-tail	0.92775164	
t Critical two-tail	1.98137181	

Table A21. Results of t-Test between the learning-based tests from hamet e and GATRAS.

Two-Sample t-Test Assuming Unequal Variances		
	CogState	GTS
Average	63.5	71.175
Variance	1618.15789	466.068289
Observations	20	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	29	
t-Statistic	-0.75183226	
P(T<=t) one-tail	0.22910539	
t Critical one-tail	1.31143365	
P(T<=t) two-tail	0.45821079	
t Critical two-tail	1.69912703	

Table A22. Results of t-Test between the learning-based tests from CogState and the results from the Ground Truth Study.

Two-Sample t-Test Assuming Unequal Variances		
	hamet e	GTS
Average	68.4525	71.175
Variance	1511.45417	466.068289
Observations	80	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	54	
t-Statistic	-0.41911114	
P(T<=t) one-tail	0.33839872	
t Critical one-tail	1.29742649	
P(T<=t) two-tail	0.67679744	
t Critical two-tail	1.67356491	

Table A23. Results of t-Test between the learning-based tests from hamet e and the results from the Ground Truth Study.

Two-Sample t-Test Assuming Unequal Variances		
	GATRAS	GTS
Average	69.1333333	71.175
Variance	2233.94802	466.068289
Observations	60	20
Hypothesized Mean Difference	0	
Degrees of Freedom (df)	70	
t-Statistic	-0.2624088	
P(T<=t) one-tail	0.3968885	
t Critical one-tail	1.2937629	
P(T<=t) two-tail	0.79377701	
t Critical two-tail	1.66691448	

Table A24. Results of t-Test between the learning-based tests from GATRAS and the results from the Ground Truth Study.

REFERENCES

1. "World Report on Disability". Report. World Health Organization. Malta. 2011.
2. "Equality Act 2010". Legislation. Crown. United Kingdom. 2010.
3. "World Economic Outlook: Legacies, Clouds, Uncertainties". Survey. International Monetary Fund. Washington. October 2014.
4. "2009 Ageing Report – Economic and budgetary projections for the EU-27 Member States (2008-2060)" Report. European Communities. Luxembourg. 2009.
5. J. Nielsen. *Usability 101: Introduction to Usability*. Nielsen Norman Group. 2012. Retrieved 26.02.2015 at <<http://www.nngroup.com/articles/usability-101-introduction-to-usability/>>
6. Oxford Dictionaries. *Definition of gamification in English*. Retrieved 23.02.2015 at <<http://www.oxforddictionaries.com/definition/english/gamification?q=gamification>>
7. "Gamification 101: An Introduction to the Use of Game Dynamics to Influence Behavior". Whitepaper. Bunchball Inc. 2010.
8. The Travel Insider LLC. *A History of US Airline Deregulation*. Retrieved 26.02. 2015 at <http://thetravelinsider.info/airlinemismanagement/airline_deregulation2.htm>
9. Entertainment Software Association. *Industry Facts*. Retrieved 26.02. 2015 at <<http://www.theesa.com/facts/index.asp>>
10. K. Kapp. "The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education". Book. John Wiley & Sons. San Francisco. 2012.
11. A. Collie, D. Darby, P. Maruff. "Computerised cognitive assessment of athletes with sports related head injury." *British journal of sports medicine* 35.5 (2001): 297-302.
12. A. A. Rizzo, J. G. Buckwalter. "Virtual reality and cognitive assessment and rehabilitation: the state of the art." *Studies in health technology and informatics* (1997): 123-146.
13. "CogState Research Manual – Version 6". Manual. CogState Limited. 2011.
14. "CogState Task Descriptions". White paper. CogState Limited.
15. "Hamet". White paper. Trainsition!
16. Berufsbildungswerk Waiblingen.. M. Dieterich, M. Goll, G. Pfeiffer, U. Weiss. *Handlungsorientiertes Testverfahren zur Erfassung und Förderung elementarer Kompetenzen für berufliche Bildung und Arbeit*. Retrieved 27. Feb 2015 at <<http://www.hamet.de/hamet-e.3109.0.html>>
17. "Analysis Guidelines". Manual. CogState Limited.