Breindebaas: measuring the Validity of a Personalised and Standardised Alcohol-Approach Bias Measure

Niels Brouwer

July 11, 2019

Supervisors: Marloes G. Postel, PhD, Melissa C. Laurens, MSc, & Marcel E. Pieterse, PhD
University of Twente, Faculty of Behavioural, Management and Social Sciences, Department of Psychology, Health & Technology, Enschede, the Netherlands

Please cite as: Brouwer, N (2019). Breindebaas: measuring the Validity of a Personalised and Standardised Alcohol-Approach Bias Measure. *Bachelor dissertation, July 2019, University of Twente, Enschede*

BREINDEBAAS: MEASURING THE VALIDITY OF A PERSONALISED AND STANDARDISED ALCOHOL-APPROACH BIAS MEASURE

Abstract

Background. Excessive alcohol consumption forms one of the leading causes of disability and mortality in the Netherlands. Dual-process models highlight the importance of the impulsive pathway in maintaining excessive drinking behaviour through the development of alcohol-approach biases. The Alcohol-Approach Bias Measure (AABM) of the mobile application *Breindebaas* may be used to measure such alcohol-approach biases. The AABM can be conducted with either standardised images (standardised AABM) or personalised images (personalised AABM). The standardised and personalised AABMs have yet to be tested on their psychometric properties. Therefore, the current study aimed to assess the validity of the standardised and personalised AABMs included in *Breindebaas*.

Methods. The current study made use of a cross-sectional design, in which seventeen participants completed the standardised and personalised AABMs alongside measures of alcohol consumption, dependence and craving, as well as a neuroticism-related measure to test for convergence. In addition, it was tested whether the inclusion of personalisation-options led to higher alcohol-approach bias scores on the personalised AABM over the standardised AABM. Lastly, the current study identified whether alcohol-approach bias scores differed on the AABMs due to moderation by one's level of neuroticism.

Results. The current study could not find support for the main study questions. Indications were provided for the insignificant findings by the feedback of participants. Participants found the minimum-number of fifty alcoholic images to be selected in the personalised AABM too large. As a result, the measure was experienced as unpersonal. Alternatively, participants believed that including multiple images of alcoholic drinks would solve the lack of personalisation, while maintaining their engagement in the conduct of the personalised AABM.

Conclusions. Despite the lack of significant findings, important implications could be drawn for the future development of *Breindebaas*. The current study contributed to the question surrounding the optimal balance between diversity in images and the experience of personalisation. Future research on personalised AABMs may remove the minimum-number of images to be selected, and include multiple images of the same drinks to evaluate the change in experience of users, and whether there occurs a corresponding change in alcohol-approach bias scores.

Excessive alcohol consumption forms one of the major causes of disability and mortality in the Netherlands. Excessive alcohol consumption can be formulated as the consumption of alcohol above recommended guidelines according to the Centers for Disease Control and Prevention (CDC, n.D.). This entails consuming more than fifteen drinks per week for men, and seven for women (CDC, n.D.). On a yearly basis excessive alcohol consumption causes 3.3 million deaths (5.3% of the deaths globally), and 132.6 million disability-adjusted life years globally (WHO, 2018). Moreover, excessive alcohol consumption is found to cause a multitude of physical ailments, including forms of head and neck cancer (Maasland et al., 2014), colorectal cancer (Owusu, Quinn, & Sheng Wang, 2014), liver cirrhosis, acute alcohol hepatitis (Adang, Wensing, & Stockbrügger, 2009), and possibly bladder cancer (Zeegers et al., 2001). Resultingly, the average amount of life years lost among the Dutch excessive drinkers equals three (WHO, 2014). Thus, it can be stated that excessive alcohol consumption forms a risk to public health in the Netherlands. The use of preventive methods is of importance to reduce the impact of this risk behaviour.

Dual-process models (of alcohol addiction) explain how alcoholism is maintained. The models propose that two neurological pathways are involved. The pathways are commonly referred to as the 'impulsive' or 'associative' pathway, and the 'reflective' pathway (Carbia, Corral, Doallo, & Caamaño-Isorna, 2018; Moss & Albery, 2009; Wiers et al., 2010[a]). Specifically, the reflective pathway is associated with higher-order evaluative processes, and is involved with emotion regulation and decision-making (Moss & Albery, 2009; Wiers et al., 2010[a]). The affective pathway is involved with covert cognitive processing, and emphasizes on the feeling of 'wanting' or 'needing' (Carbia et al., 2018; Moss & Albery, 2009).

Expectedly, functional differences in the neurological pathways were found as a result of excessive alcohol consumption. In healthy individuals the pathways interact to balance higher-order goals with lower-order wants (Boffo, Pronk, Wiers, & Mannarini, 2015). In contrast, it is argued that engagement in excessive alcohol consumption causes the reflective pathway to be hyporesponsive, and the affective pathway to be hyperresponsive (Moss & Albery, 2009). Consequentially, excessive alcohol consumption leads to reduced cognitive control, whereas an increased feeling of 'wanting' to consume alcohol causes one to be more likely to drink (Carbia et al., 2018). Therefore, if one wants to successfully decrease the amount of alcohol consumed through interventions, both pathways have to be accounted for (Heitmann et al., 2017).

To address both pathways, the alcohol-approach bias is of importance to manipulate

(Eberl et al., 2013; Heitmann et al., 2017). The term 'alcohol-approach bias' can be defined as automatically generated action tendencies caused by substance-related cues in the environment (Manning et al., 2016; Wiers et al., 2010[a]). Alcohol-approach biases are formed through processes of associative learning and operant conditioning (Boffo et al., 2015). As these forms of learning occur outside of consciousness, alcohol-approach biases are often unknown to the individual, and difficult to address through treatment (Boffo et al., 2015).

Fortunately, techniques were developed that may be used to measure alcohol-approach biases in individuals. One such technique is the Relevant Approach-Avoidance Task (R-AAT) (Wiers et al., 2010 [b]). The R-AAT contains two randomly ordered trials in which alcoholic and non-alcoholic images have to be interacted with via computer. An alcohol-approach response is simulated by combining the view of alcohol-related images with an action that resembles a movement towards the person, such as pressing the backward-arrow-button on a keyboard (Eberl et al., 2013; Manning et al, 2016). In contrast, an alcohol-avoidance response is simulated by combining the images with a movement away from the person, such as pressing the forward-arrow-button (Manning et al., 2016; Wiers et al., 2010[b]). In the second trial of the R-AAT, the push-and-pull conditions are reversed. The resulting measure of the strength of the alcohol-approach bias is commonly referred to as an Alcohol-Approach Bias measure (AABM). AABMs are established by identifying one's response times (RTs); the stronger one's bias, the slower one will push away the alcoholic drink, and the faster one will pull the alcoholic drink towards oneself (Wiers et al., 2010 [b]).

In an attempt to compare the quality of AABMs, Kersbergen, Woud, and Field (2015) contrasted the psychometric properties of the R-AAT with the Relevant Stimulus-Response Compatibility task (R-SCR). The researchers made use of computerised versions of the AABMs. Scores on the R-AAT and the R-SCR were found to positively correlate with scores on the Alcohol Use Disorders Identification Test (AUDIT) (Babor et al., 2001). In other words, both measures accurately predicted hazardous drinking (Kersbergen et al., 2015). Interestingly, the R-AAT was the sole measure positively correlating with the Alcohol Timeline Followback (Alcohol TLFB) (Sobell & Sobell, 1995). This indicates that the R-AAT is able to accurately predict (weekly) alcohol consumption. As a result of these findings, it may be stated that the R-AAT forms an accurate measuring instrument in regard to alcohol consumption and alcohol-bias levels, if not one of the more accurate AABMs available.

The current relevancy of M-Health (Mobile Healthcare), and its associated use of smartphones, has led to the translation of the R-AAT to the smartphone-platform (Somsen,

2017). The mobile translation of the R-AAT is called *Breindebaas* [Being the boss of your own brain] (Tactus Verslavingszorg, n.D.). *Breindebaas* includes two different options: the option for measuring alcohol-approach biases; the AABM, as well as trainings to reduce one's alcohol-approach bias over an extended period of time; the Alcohol-Approach Bias Training (AABT). The current study emphasises on the AABM included in *Breindebaas*. As of current, research has not been conducted in regard to the AABM. In contrast, the AABT was evaluated on its efficiency in reducing alcohol-intake levels in participants. In specific, Somsen (2017) and Nijen Es (2017) found indications of the AABT being effective in reducing alcohol-consumption levels up to four months follow-up.

Additional feedback retrieved from the AABT-pilot-studies' participants included a need for personalisation of the alcoholic and non-alcoholic images. Personalisation-options were added for the AABT as a result. Currently, the question arose whether to include personalisation-options in the AABM as well, and how this could best be implemented. Recently, the option to personalise images was included in a more recent version of *Breindebaas* (Postel, personal communication, March 7, 2019). Hence, there now exists two AABMs in *Breindebaas*; the standardised AABM and personalised AABM. Interestingly, the current study found that the implementation of personalisation-options showed to link to the incentive-sensitisation theory of addiction (Robinson & Berridge, 2008). The theory proposes that biased cognitive processes are highly dependent on the association individuals developed between specific stimuli and its following rewarding experience. In other words, the AABM may be more accurate in identifying an alcohol-approach bias when personalised images are used, as the content of these images is more relevant to the individual.

As of current, the psychometric properties of the standardised and personalised AABMs in *Breindebaas* are yet to be identified. Hence, the current study will focus itself on the convergent validity of the AABMs. As stated, the standardised AABM included in *Breindebaas* is a mobile translation of the R-AAT provided by Wiers and others (2010[b]). Hence it is expected that the standardised AABM knows similar, if not equal psychometric properties to the R-AAT, as found by Kersbergen and others (2015). The standardised AABM may therefore be therefore be taken as the golden standard in assessing the psychometric properties of the standardised and personalised AABMs. Due to the expected similarities between the psychometric properties of the standardised AABM and the R-AAT, the current study shall also test the convergence between the AABMs and the alcohol-related measures used by Kersbergen and others (2015) in assessing the R-AAT. The measures include the

AUDIT (Babor et al., 2001) and the Alcohol TLFB (Sobell & Sobell, 1995). In addition, a Craving Visual Analog Scale (Craving VAS) was added as an alcohol-related measure. Lastly, the current study will assess whether the personalised AABM measures stronger alcohol-approach biases than the standardised AABM, in accordance with the incentive-sensitisation theory of addiction (Robinson & Berridge, 2008). In other words, the current study will identify whether alcohol-approach bias scores increase when images are more personally relevant.

Neuroticism

The personality trait 'neuroticism' forms a relevant topic for *Breindebaas*. Neuroticism is defined as a predisposition to experience greater levels of emotional distress and negative affect (Papachristou, Nederkoorn & Jansen, 2016). Highly neurotic individuals are found to form an at-risk group in developing stronger alcohol-approach biases. According to the tension reduction hypothesis, highly neurotic individuals may consume more alcohol to reduce their stress levels. Resultingly, neurotic individuals may be at a greater risk for excessive alcohol consumption, as their levels of distress tend to be more impactful (Liu, Wang, Zhan, & Shi, 2009). This hypothesis has been confirmed in different situations, ranging from coping with work stress (Liu et al., 2009) to coping with childhood traumas (Schwandt et al., 2013). Secondly, it is claimed that the negative reinforcement due to alcohol consumption in stressful situations (i.e. stress reduction) may lead alcohol to be evaluated as relaxing and supporting. Positive evaluations may result in positive arousal states and pro-alcohol attitudes towards alcoholic beverages (Salemink et al., 2015; Lac & Donaldson, 2016).

In addition to a greater likelihood to develop alcohol-approach biases, highly neurotic individuals may also be more influenceable by personalisation-options in mobile applications (Liu et al., 2009; Salemink et al., 2015). In a (simulated) attempt to adjust participants' transportation habits to be in favour of eco-friendly options, Anagnostopoulou and others (2017) tempted to identify the relationship between personality types and the most suitable persuasive strategies. Specifically, the authors found high levels of neuroticism to relate strongest with the strategy 'personalisation'. Indications are not provided as to why the relationship exists. Nevertheless, the finding may suggest that highly neurotic individuals experience a greater need for the content to be more relatable to the self as to increase its effect. Further confirming the finding may have great implications for the field of excessive

alcohol consumption, as it would provide guidelines on how to make *Breindebaas* more efficient for one of its at-risk groups.

From the above-mentioned findings, two important implications can be identified. Firstly, it may be that high levels of neuroticism relate to a greater likelihood of experiencing an alcohol-approach bias. Hence, neuroticism-related measures alike the Big Five Inventory – Neuroticism subscale (BFI – N) (John & Srivastava, 1999) may converge significantly with the AABMs of *Breindebaas*. Secondly, a high level of neuroticism may lead to differences in alcohol-approach bias scores on the standardised versus personalised AABMs, as highly neurotic individuals are found to experience greater affect by personalisation-options.

Study questions

In regard to the validity of the standardised and personalised AABMs in *Breindebaas*, the current study will tempt to address the following study questions:

(1) *'Is there a significant correlation between the personalised AABM and the standardised AABM of Breindebaas?*'. As the personalised and standardised measures are identical aside of the option to personalise images, it is expected that a significant, positive correlation will be found.

(2) '*Are the correlations between the standardised and personalised AABMs, and the Alcohol TLFB, AUDIT, and the Craving measure significant?*'. Due to the positive correlations found between the R-AAT, and the Alcohol TLFB and AUDIT (Kersbergen et al., 2015), it is expected that similar results will be found for *Breindebaas*. In addition, a positive correlation between the AABMs and the Craving VAS is expected due to the shared construct of both measures (i.e. alcohol-approach bias). Lastly, the correlations are expected to be higher for the personalised AABM than the standardised AABM, in accordance with the incentive-sensitisation theory of addiction (Robinson & Berridge, 2008).

(3) '*Are the alcohol-approach bias scores obtained on the personalised AABM significantly higher than the standardised AABM*?'. It is expected that the hypothesis will be confirmed, in line with the previously explained incentive-sensitisation theory of addiction (Robinson & Berridge, 2008).

Moreover, the identified relationship between alcohol-approach biases and neuroticism, and the possibility of an enhanced effect of personalisation-options on highly neurotic individuals, have led to the following study questions:

(4) '*Are there significant correlations between the standardised and personalised AABMs with the BFI- Neuroticism Subscale?*'. A significant positive correlation is expected, in accordance with the tension reduction hypothesis (Liu, Wang, Zhan, & Shi, 2009).

(5) '*Does neuroticism moderate the relationship between scores on the standardised AABM and scores on the personalised AABM?'.* It is expected that a high score on the BFI-N may lead to a lower correlation between the standardised and personalised AABM. Specifically, the lower correlation is believed to exist by higher scores than the low-neuroticism group on the personalised AABM, and lower scores than the low-neuroticism group on the standardised AABM. This is, as highly neurotic individuals were found to have a greater need for personalisation-options, and thereby may depend more on self-relevant images to identify an alcohol-approach bias (Anagnostopoulou et al. 2017). In contrast, the difference in scores on the standardised and personalised AABM for low-neurotic individuals is believed to be smaller.

Methods

Participants

Participants had to meet two inclusion criteria: (1) participants were required to be at least eighteen years old, and (2) males had to drink equal to or more than seven drinks per week, and females equal to or more than four drinks per week. The values were established by halving the norms for excessive alcohol consumption (CDC, n.D).

Participants were recruited with the use of a convenience sampling procedure through a multitude of networks. The networks included personal networks, third-party networks (through family and friends), and social media networks (e.g. Facebook; Snapchat). A flyer was spread at multiple locations in Almelo and Enschede. The locations include local libraries, supermarkets, and community centres. The flyer was also posted online through social media (e.g. Facebook; Whatsapp). Lastly, the flyer was promoted on the local news channel of Almelo (AAVisie), via their website and television channel.

In total, 27 participants were recruited. Of the 27 participants, eight participants did not finish the AABM-measures (fully), and nine participants did not finish the questionnaire (fully). Note that two participants withdrew from the study previous to the conduct of the AABMs, and hence solely finished the questionnaire. Moreover, responses were excluded based on the AAT-guidelines proposed by Nosek and others (2014). This led the AABM-measures of two participants to be deemed unusable due to 25% of their responses (1) having

a reaction time above 2000ms, or (2) being erroneous. An interesting side-note is that the excluded participants were also the oldest participants in the current sample (age ≥ 65). Ultimately, ten participants were excluded from the data. The resulting dataset included seventeen participants with a mean age of 41.53 (*SD* = 14.04), ranging from 22 to 64 years old. Twelve participants were males, and five participants were females.

Design

The current study deployed a cross-sectional design. Participants were first required to fill in a questionnaire. Followingly, the participants conducted the standardised and personalised AABMs.

Intervention

Breindebaas app. *Breindebaas* is a mobile application developed by Tactus Verslavingszorg [Addiction treatment] (n.D.), in cooperation with the University of Amsterdam, University of Twente, and the Saxion University of Applied Sciences. Although the app is still in development as of current, it will purportedly be used to (a) measure alcohol-approach biases in individuals (AABM), and/or (b) train individuals in reducing their alcohol-approach bias (AABT). Notably, the current research focusses on the AABM of *Breindebaas*. Moreover, the current study shall test the second version of *Breindebaas*. The second version includes the options to either use standardised images, or to personalise the images that will be displayed during the measure. The second version of the app is still in development, and hence not yet provided over app-stores (e.g. Play Store; App store). The app is provided for solely a Dutch audience; this accounts for the language used, as well as the images included.

Measurements

AABMs. The current study made use of the standardised and personalised AABMs in *Breindebaas*. The conduct of the AABMs involved two phases in which 40 images (excluding ten practice-images) of alcoholic and non-alcoholic drinks had to be swiped towards oneself or away, based on the instructions provided by the mobile application. During the second phase, the swipe movements were reversed for the images. The movements were supplied by corresponding sounds for correct and incorrect answers.

Additionally, the current study made use of an online questionnaire, provided through Qualtrics (Qualtrics, Provo, UT). The following questions and/or measurement instruments were included:

Baseline characteristics. The questionnaire included the baseline characteristics: age, gender, primary daily activity (e.g. studying/working), and the highest achieved academic degree.

Alcohol Timeline Followback (Alcohol TLFB). The Alcohol TLFB, developed by Sobell and Sobell (1995), is a measurement instrument used to determine the amount of daily drinking of individuals. In the current study, the questions were limited to the participants' consumption levels over the past week. Psychometric properties of the Alcohol TLFB were found to be adequate. Namely, the Alcohol TLFB showed high test-retest reliability amongst low and high-problem drinkers, and moderate male and female drinkers taken from the general population (Sobell & Sobell, 1992). Moreover, scores on the Alcohol TLFB were found to correlate significantly with the total number of drinks consumed by participants, the total amount of drinks consumed divided by the total amount of drinking days, as well as abstinence days (Sobell, Brown, Leo, & Sobell, 1996). The Alcohol TLFB does not know norm tables. Hence, participants were classified as 'high' or 'low' in their weekly alcohol consumption levels by identifying whether they scored above or below the sample mean.

Alcohol Use Disorders Identification Test (AUDIT). The AUDIT is a measurement instrument developed by Babor and others (2001). In the current study, a Dutch translation of the AUDIT, provided by Schippers and Broekman (2010) was used. The AUDIT is used to screen for problematic drinking behaviour and alcohol dependency. The AUDIT includes nine questions regarding alcohol intake levels and related factors to alcohol consumption. Specifically, the first seven questions are placed on a five-point-scale, and are scored from zero to four points. The remaining two questions, which regard injuries and received advice to lessen alcohol intake by others, are placed on a three-point-scale with the responses being scored as zero, two, or four. The psychometric properties of the AUDIT are satisfactory. Test-retest reliability was found to range between .80 and .84 (Kim, Gulick, Nam, & Kim, 2008; Meneses-Gaya, Zuardi, Loureiro, & Crippa, 2009). The AUDIT also knows high construct validity, showing an alpha-value of .94 (Meneses-Gaya et al., 2009). A scoring form is provided for the AUDIT (Australian Government, Department of Veterans' Affairs, n.D.). The scoring form was used to label individual participants, as well as the sample mean.

Craving Visual Analog Scale (Craving VAS). The Craving VAS is a self-developed measure, which determines the current level of alcohol priming in individuals. The Craving VAS was placed on a ten-point-scale, with the value '0' standing for no current craving for alcohol, and '10' standing for a high current craving for alcohol. Participants were labelled 'high' or 'low' based on their VAS-scores being above or below the sample mean.

Big Five Inventory – Neuroticism subscale (BFI–N). The Big Five Inventory (BFI) (John & Srivastava, 1999) is a self-report personality questionnaire, consisting of five different personality domains: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. The BFI was developed based on the Big Five Factors of personality, as identified by Goldberg (1992; 1993). The BFI requires one to answer 44 questions based on a five-point Likert scale. The scores on this scale are labelled: (1) disagree, (2) slightly disagree, (3) neutral, (4) slightly agree, and (5) agree. The current study solely made use of the subscale 'Neuroticism' (BFI–N). The BFI-N includes nine questions. The subscale was found to uphold adequate psychometric properties when used as a stand-alone measure. Reliability of the BFI-N was found to range between .81 and .86 (Denissen et al., 2008; Leung et al., 2012). Moreover, the scale's validity showed to be .68 (Denissen et al., 2008). The BFI does not know official norm tables. Hence, participants 'high' or 'low' in neuroticism were determined by identifying whether these scored below or above the mean. Notably, the aforementioned approach is recommended by the developer of the BFI (Srivastava, 2012).

Procedure

Research procedure. The following section will describe the specific steps of the study. The standardised text used for the study can be found in Appendix 2, and shall be referred to throughout the section. The research procedure was conducted as follows.

Before the conduct of the study, participants were randomly assigned to the first measurement condition. For this, the program 'Random Team Generator' was used (Random Lists, n.D.) to assign participants to the 'standardised' or 'personalised' AABM-group. Herein, the participants were identified by number. The numbers resembled the chronology of when the participants would be tested.

At the start of the research, participants were thanked for their interest in participation. After this, they were asked to fill in the online questionnaire on a phone supplied by the researchers. The online questionnaire started with an informed consent, which the participants

had to agree with before they could continue with the questionnaire and study (Appendix 2: Section 1.1).

After this the app was used. The participant had to register oneself. Followingly, the participant was asked to log into the app with their account details. After logging in, a screen popped up that showed a multitude of options. During this, the researcher made sure to activate the options 'Activate measures' and 'Play a sound with right or wrong answers' (Appendix 2: Section 1.2). Moreover, this screen included the list of the alcoholic and non-alcoholic drinks one prefers. When participants started in the personalised condition, the list of images had to be adjusted (Appendix 2: Section 2.1).

After the selection of images, the participant was asked to press the 'Start session' button, and follow the instructions provided by the app. During the session, the researcher wrote down comments regarding malfunctions of the app (e.g. the app got stuck during the measurement), and difficulties during the studies external of the app (e.g. distractions).

After finishing the first AABM, the researcher logged out the account, and closed the app fully as to secure the data. Then the break started. During the break, the participant was asked whether one would like to finish a puzzle. If the participant would prefer not to finish a puzzle, one was allowed to interact with their phone (Appendix 2: Section 2.3).

The second AABM was taken after the break. The participant was asked to log back in on the app and continue with the second measurement. Now, the participant had to conduct either the generalized or personalised AABM of the app, based on what option was left over (Appendix 2: Section 2.1; 2.2).

When the second measurement was finished, the research came to an end. The researcher made sure to log out the account and close the app completely again. The researcher thanked him or her for one's time and participation, and asked about the experiences of the participant during the research (Appendix 2: Section 3).

Data analysis

Participants' characteristics. For the data analysis, IBM SPSS Statistics 25 was used. Firstly, descriptive statistics were used to identify the sample's characteristics. Characteristics included the variables: sex, age, primary daily activity, and the highest obtained educational degree. Furthermore, the d-score on alcohol of the sample was calculated. In specific, the d-scores on alcohol translate into the strength of the alcohol-approach bias in participants. The d-scores were derived from both the personalised AABM

and the standardised AABM. The d-scores were measured by subtracting the mean reaction time in the alcohol-pull-trials from the mean reaction-time in the alcohol-push-trials, and dividing the outcome by the pooled standard deviation of alcohol: $(M_{RT\text{-}push\text{-}alcohol} - M_{RT\text{-}pull\text{-}alcohol})/SD_{alcohol}$ (Greenwald, Nosek, & Banaji, 2003; Nosek et al., 2014). D-scores range from -2 to 2. The resulting value may indicate an approach bias to alcohol when it is positive. In contrast, a negative value indicates an avoidance bias to alcohol.

Convergent validity. To answer the first study question, a correlation analysis was conducted to determine the relationship between the variables 'd-score standardised AABM' and 'd-score personalised AABM'. Previous to answering the second study question, baseline characteristics of the sample were identified for the variables: mean Alcohol TLFB scores, mean AUDIT scores, and mean Craving VAS scores. Followingly, the second study question was answered by using a multiple correlation analysis to establish the correlation between the standardised and personalised AABMs individually with the Alcohol TLFB, AUDIT and Craving VAS scores. Thus, the variables included were the 'd-score standardised AABM' and the 'd-score personalised AABM', which were individually correlated with the variables 'Alcohol TLFB score', 'AUDIT score', and 'Craving VAS score'.

Effect of personalisation on measurement accuracy. For the third study question, the current study conducted a dependent samples t-test to identify whether the difference in mean d-scores between the variables 'd-score standardised AABM' and 'd-score personalised AABM' were significant.

Neuroticism. Normality of neuroticism levels in participants was identified with Kolmogorov-Smirnov and Shapiro-Wilk tests. Outliers in the variable 'neuroticism' were identified by using Tukey's method, which identifies extreme values falling outside an IQR of 1.5. Baseline differences for the subsamples 'high' and 'low' in neuroticism were identified in regard to their age, gender, primary daily activities, achieved academical degrees, and mean Alcohol TLFB scores, AUDIT scores and Craving VAS scores. Followingly, the fourth study question was approached with univariate correlation analyses, as to determine the convergent validity between the standardised and personalised AABMs with the BFI-N. The variables 'd-score standardised AABM', 'd-score personalised AABM' and 'BFI-N score' were used.
To answer the fifth research question, a Baron and Kenny (1986) moderation analysis was conducted to identify whether neuroticism moderates the association between d-scores on the standardised AABM and d-scores on the personalised AABM. To examine for moderation, a hierarchical multiple linear regression was conducted. The independent variables of the

regression include ‘d-score standardised AABM’ and ‘BFI-N score’, followed by the interaction between ‘d-score standardised AABM’ and ‘BFI-N score’. The interaction term was created by multiplying the variables ‘d-score standardised AABM’ and ‘BFI-N score’. The dependent variable of the regression forms the ‘d-score personalised AABM’. Moderation by neuroticism is confirmed when the interaction is significant.

Feedback by participants on Breindebaas. Feedback provided by participants was jotted down on paper at the end of the research-session. Key words from the points of feedback were identified as to create overlapping statements. The statements were labelled as either ‘Positive’, ‘Neutral’, or ‘Negative’ depending on their evaluation by participants. Ultimately, the number of participants whose feedback points connected to the statements were noted down.

Results

Participants’ characteristics

The sample previous to exclusions contained 27 participants. The sample had a mean age of 40.7 (*SD* = 16.58), with the youngest participant being 19, and the oldest being 79. The sample existed of 70.4% males and 29.6% females. Primary daily activity was being a student for 25.9% (*n* = 7) of the sample, 63% (*n* = 17) had a paid occupation, and 11% (*n* = 3) indicated to have different daily activities. The different activities included a ‘working disability’, ‘voluntary work’ and ‘retired’. The highest obtained educational degree by participants was elementary school or a practical education for 11.1% (*n* = 3) of the sample, a high school degree or an intermediate vocational educational degree for 48.1% (*n* = 13) of the sample, and a degree at a university of applied sciences or a university for 37% (*n* = 10) of the sample. One participant (3.8%) was found to have a different educational background, which included an ‘MHNO’ degree.

Participants after exclusions (*n* = 17) had a mean age of 41.53 (*SD* = 14.04) and were mainly males (*n* = 12, 70.6%). In regard to one’s primary daily activity, 17.6% of the participants (*n* = 3) showed to be students, and 70.6% (*n* = 12) had a paid occupation. Two individuals (11.8%) had other primary daily activities, which included ‘working disability’ and ‘voluntary work’. The highest obtained educational degree showed to be elementary school or a practical education for two participants (11.8%), a high school or an intermediate vocational educational degree for seven participants (41.2%), and a degree in a university of

applied sciences or a university for seven participants (41.2%). One participant (5.9%) obtained a different degree, which was an 'MHNO' degree.

Followingly, the d-scores of the participants were calculated (Appendix 3: Table 1; Table 2). Means for the sample were .18 (*SD* = .49) on the personalised AABM, and .20 (*SD* = .52) on the standardised AABM. Notably, the mean d-scores show a modest positive alcohol-approach bias in participants on both AABMs. Normality checks of the participants' d-scores on the standardised and personalised AABMs were conducted. D-scores on the standardised and personalised AABMs were normally distributed according to the Kolmogorov-Smirnov test (D = .126, p = .20; D = .228, p = .19) and Shapiro-Wilk test (D = .920, p = .15; D = .968, p = .78).

Convergent validity

A univariate correlation analysis was conducted to correlate the d-scores on the personalised AABM with the d-scores on the standardised AABM. Results of the Pearson correlation showed that the correlation between d-scores on the personalised AABM and standardised AABM was remarkably low, $r(15)$ = .009, p = .974. As the correlation between the personalised AABM and standardised AABM shows to be non-significant, the first hypothesis was refuted. Nevertheless, a scatterplot with fit line (Figure 1) was used to further analyse the low correlation between the standardised and personalised AABMs.

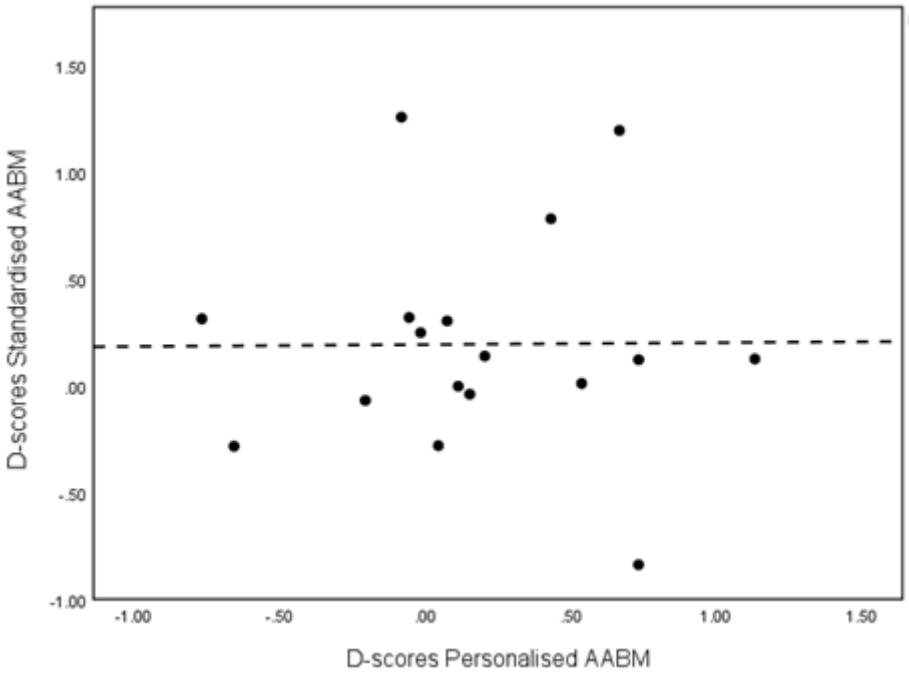

Figure 1. Correlation between d-scores on the standardised and personalised AABMs taken by 17 participants. Pearson's *r* = .009

The spread between data points in the scatterplot shows to be large, leading the fit line to be approximately horizontal. Notably, d-scores of four participants on the standardised and personalised AABMs were found to differ substantially compared to the overall sample. Specifically, case numbers 1, 5, 12, and 14 obtained d-scores on the standardised and personalised AABMs with a score difference of equal to or more than .75 (see Appendix 3: Table 1; Table 2). Tukey's method was used with the variables 'd-score standardised AABM' and 'd-score personalised AABM' to further identify outliers within the sample outside of an IQR of 1.5. For the d-scores on the standardised AABM, Tukey's method confirmed that case number 5 surpassed the upper bound of the sample (d-score = 1.26), whereas case number 12 (d-score = -.84) surpassed the lower bound of the sample. No indications were provided regarding case numbers 1 and 14. Regardless, the four participants with substantial d-score differences were excluded from the data, after which a univariate correlation analysis was run with the remaining data of 13 participants.

The Pearson correlation revealed that the correlation between d-scores on the standardised and personalised AABMs with the exclusion of the four participants was significant, $r(11) = .569$, $p = .042$. A scatterplot with fit line (Figure 2) was created to further observe the relationship between d-scores on the standardised and personalised AABMs. The scatterplot provides indications that higher d-scores obtained on the personalised AABM correspond with higher d-scores obtained on the standardised AABM for circa 70% of the participants.

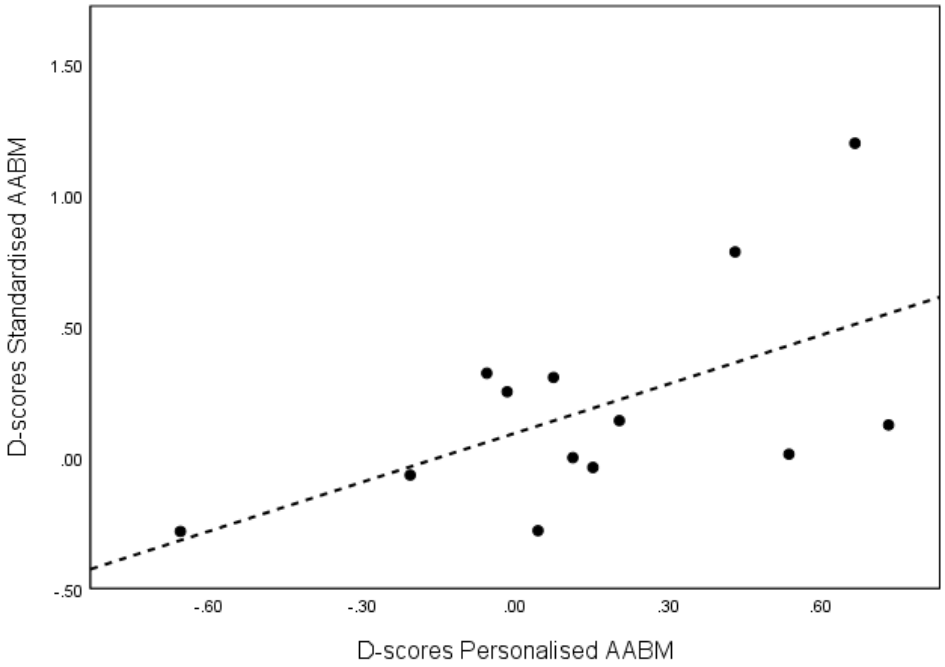

Figure 2. Correlation between d-scores on the standardised and personalised AABMs taken by 13 participants. Four participants were excluded due to substantial differences in d-scores on the AABMs (difference ≥ .75). Pearson's *r* = .569

Note, however, that the first hypothesis shall not accepted as a result of this finding. Namely, two of the four outliers were not possible to be identified by using outlier detection techniques such as Tukey's method. Additionally, participants with erroneous data in d-scores on the AABMs were previously excluded based on the AAT-guidelines (Nosek et al., 2014). Hence it is likely that the current participants were excluded based on other reasons than erroneous data, such as difficulties experienced with the AABMs, or simple random variation.

Furthermore, descriptive statistics were established for scores on the alcohol-related measures. The mean scores were 10.89 (*SD* = 4.58) on the Alcohol TLFB, 8.47 (*SD* = 3.61) on the AUDIT, and 3.55 (*SD* = 2.80) on the Craving VAS. In regard to the norm tables for the AUDIT, the sample was labelled as 'Moderate'. In other words, the sample knows a moderate risk of physical or psychological harm as a result of their alcohol consumption behaviour. Participants' individual scores and labels on the alcohol-related measures can be found in Appendix 4: Table 3.

Followingly, a multiple correlation analysis was used to calculate the convergence between the d-scores of the standardised and personalised AABMs with Alcohol TLFB, AUDIT, and Craving VAS scores. Results of the multiple correlation analysis can be found in Table 1. The multiple correlation analysis showed no significant correlations between d-

scores on the two AABMs with scores on the Alcohol TLFB, AUDIT and the Craving VAS. Ultimately, the second hypothesis was refuted.

Table 1

Correlations of the D-scores on the Personalised and Standardised AABM with the Alcohol TLFB, AUDIT, and Craving VAS scores (N = 17)

AABM	*r*	*p*
Personalised AABM		
Alcohol TLFB	.229	.376
AUDIT	.034	.898
Craving VAS	-.135	.605
Standardised AABM		
Alcohol TLFB	-.325	.203
AUDIT	-.063	.810
Craving VAS	-.183	.481

Note. The Alcohol TLFB is developed by Sobell and Sobell (1995). The AUDIT is developed by Babor and others (2001).
Note. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Notably, the correlations between the standardised AABM and the Alcohol TLFB, AUDIT and Craving VAS all show to be negative. The finding indicates that participants who have a strong alcohol-approach bias score low on other alcohol-related measures of interest. Contrastingly, correlations between d-scores on the personalised AABM and scores on the Alcohol TLFB, AUDIT, and Craving VAS were found to be more divergent. In specific, the correlation between the personalised AABM and the Alcohol TLFB neared the current study's expectations. A scatterplot with fit line (Figure 3) was used to further analyse the relationship. The scatterplot reveals a trend to exist, in which higher d-scores on the personalised AABM pair with greater weekly alcohol consumption levels as measured with the Alcohol TLFB.

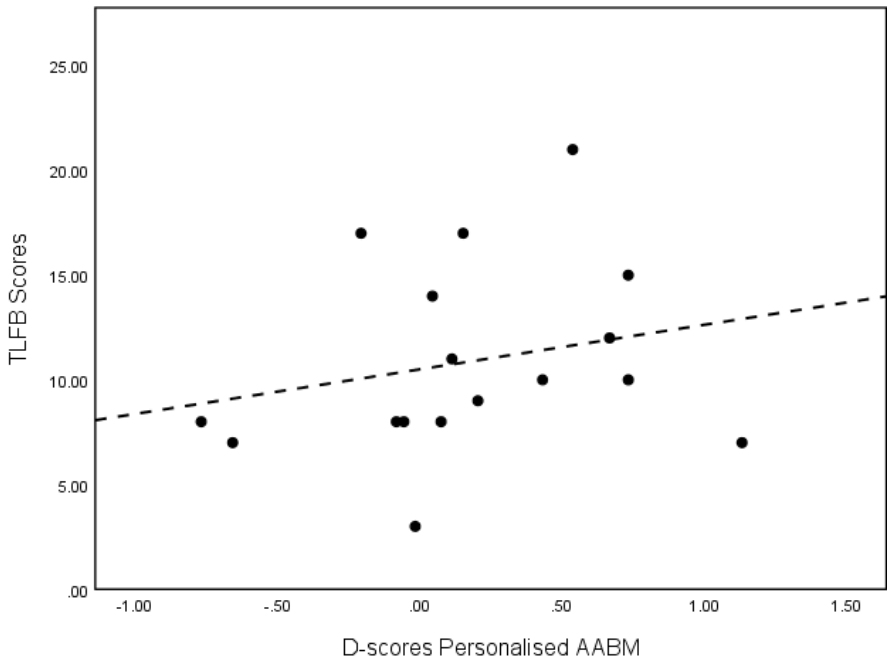

Figure 3. Relationship between scores on the Alcohol TLFB and d-scores on the personalised AABM taken by 17 participants. Pearson's r = .229

Effect of personalisation on measurement accuracy

A dependent samples t-test was used to compare d-scores of the personalised AABM and the standardised AABM. There was a non-significant difference in the d-scores of the standardised AABM (M = .20, SD = .52) and the personalised AABM (M = .18, SD = .49); t(16) = .098, p = .923. This finding indicates that d-scores on the personalised AABM are not significantly higher (or lower) than d-scores on the standardised AABM due to the addition of self-relevant images. Resultingly the third hypothesis was refuted.

Neuroticism

The sample mean on the BFI-N was 24.82 (SD = 5.99). Individual scores on the BFI-N are found in Appendix 4: Table 3. Normality checks of the BFI-N scores were established by using Kolmogorov-Smirnov and Shapiro Wilk tests. Levels of neuroticism showed to be normally distributed according to the Kolmogorov-Smirnov test, D = .204, p = .058. In contrast, the Shapiro-Wilk test indicates levels of neuroticism not to be normally distributed, D = .889, p = .045. In other words, it may be possible that participants have scored higher or lower on the BFI-N than was expected. However, as findings from the Kolmogorov-Smirnov test and Shapiro-Wilk test showed to contradict, the non-normality of the distribution on BFI-N scores was believed to be negligible. Hence, parametric tests were run despite the finding.

Subgroups 'high' and 'low' in neuroticism were tested for baseline differences in age, gender, daily activity, academical degree, mean TLFB scores, AUDIT scores, and Craving VAS scores (Table 2). The subgroups were established by grouping participants with BFI-N scores above and below the mean of the sample ($M_{BFI\text{-}N}$ = 24.82) (Appendix 4: Table 3). As can be seen in Table 2, no significant differences at baseline in all variables between participants with a high versus low neuroticism score was found.

Table 2

Descriptive Statistics on the High-Low Neuroticism Subgroups

	All participants		High Neuroticism		Low Neuroticism		
	M	*SD*	*M*	*SD*	*M*	*SD*	*p*
Age (years)	41.53	14.04	42.89	12.43	40.00	16.40	.686
Gender (%male)	70.6%		77.8%		62.5%		.490
Daily activity (label)	PO		PO		PO		.468
Academical degree (label)	HS/IVO		HS/IVO		HS/IVO		.129
Alcohol TLFB	10.89	4.58	10.33	4.55	11.50	4.84	.616
AUDIT	8.47	3.61	8.56	4.19	8.38	3.11	.922
Craving	2.29	2.82	2.00	2.87	2.63	2.92	.663

Note. N = 17. High Neuroticism (*n* = 9), Low Neuroticism (*n* = 8).
Note. 'Age', 'Alcohol TLFB', 'AUDIT' and 'Craving' were analysed with a MANOVA (F[1,15]). A chi-square test was used for 'Gender' (*df* = 1), 'Daily activity' (*df* = 1), and 'Academical degree' (*df* = 2). The *p* values are two-tailed.
Note. Categories and labels on the variable 'Daily activity' include: Student (S), Paid Occupation (PO).
Note. Categories and labels on the variable 'Academical Degree' include: Elementary School/Practical Education (ES/PE), High School/Intermediate Vocational Education (HS/IVO), University of Applied Sciences/University (UAS/U).
Note. **p* < 0.05. ***p* < 0.01. ****p* < 0.001.

To identify whether d-scores on the personalised and standardised AABMs converged with scores on the BFI-N, univariate correlation analyses were conducted. The Pearson correlation revealed that there was no significant correlation between d-scores on the personalised AABM and scores on the BFI-N, $r(15) = .123$, $p = .638$. In addition, no significant correlation between d-scores on the standardised AABM and scores on the BFI-N was found, $r(15) = -.264$, $p = .305$. As the BFI-N showed not to correlate significantly with both AABMs, the fourth hypothesis was refuted.

To test the hypothesis that the difference between d-scores on the standardised and personalised AABMs are greater for highly neurotic individuals, a Baron and Kenny (1986) moderation analysis was used. To examine for moderation, a hierarchical multiple regression analysis was deployed. The dependent variable in the following analyses was 'd-score personalised AABM'. In the first step, two independent variables were included: 'd-score standardised AABM' and 'BFI-N score'. The independent variables did not significantly

account for the amount of variance observed in the obtained d-scores on the personalised AABM, R^2 = .072, $F(2, 14) = 5.40$, p = .594. Next, the interaction between the d-scores on the standardised AABM and BFI-N scores was added to the regression model as an independent variable. The interaction did not account for the proportion of variance observed in the obtained d-scores on the personalised AABM, ΔR^2 = .00, $\Delta F(1, 13) = .002$, p = .800, b = .003, t(13) = .40, p = .969. In other words, high levels of neuroticism were not found to cause significant differences in the correlation between d-scores on the standardised and personalised AABMs, in contrast to those obtained by low-neurotic participants. No further analyses were conducted as a result of the insignificant findings. Ultimately, the fifth hypothesis was refuted.

Feedback on Breindebaas by participants

Feedback by participants can be found in Appendix 5. Table 4 sums up participants' individual feedback points regarding the *Breindebaas*-app. Table 6 includes the grouped responses of participants. Notably, the feedback includes comments made by participants who were excluded from the research, but did interact with *Breindebaas* (n = 25). Although their data was deemed unusable, it is believed that the points of feedback given on *Breindebaas* may still be of significant value to further improve the mobile application. Prominence was determined by identifying feedback points brought forward by a minimum of 10% of the participants ($n \geq 3$).

The 'Positive' label included the statement 'enjoyable game element'. This statement was brought forward by three participants. The participants indicated to see the measures as a game, in which they were opted to score as fast as possible. Mistakes were often followed by a sense of being bothered or aggravated.

The 'Neutral' label included a question shared by seven participants. The question entailed 'uncertainty whether there were tricky images, such as alcohol-free beer'. Specifically, participants hesitated most about the classification of 'Radler'. Radler is a mixture of beer and lemonade, and knows an alcoholic and non-alcoholic variant.

The 'Negative' label included eight points of feedback as to further improve *Breindebaas*. The most prominent statement was shared by 19 participants, and included that there were 'too many images to select'. In specific, 19 participants believed that the minimum number of images that had to be selected per category, which is 50, were too many. Participants indicated either not to like, or not to know 50 different drinks. The statement was

made mainly in regard to the alcoholic category. Of the 19 participants who believed there were too many images to select, ten participants indicated this to cause the personalised AABM to be 'not personalised'.

Furthermore, six participants stated that there were 'double-up images', or multiple images of the same drink. Interestingly, this contrasted the statement: 'add multiple images of one drink', which was shared by three participants. It is believed that the aforementioned statements connect with a third statement: 'categorize images per brand/type of drink', which was provided by five participants. This statement entailed that categories should be established where multiple images of the same drink could be found.

Moreover, participants made statements surrounding the clarity of the images. Three participants indicated that 'not all images were clear'. The same participants indicated that the size of images could be increased to solve this matter.

Lastly, four participants indicated that the measure creates a thirst for alcohol. In this, two participants indicated that they would consume an alcoholic beverage after the research had ended.

Notably, the head-researchers/mentors of the current research team indicated this study to be the first *Breindebaas*-study that explicitly emphasised on the feedback provided by elders ($n = 2$, age ≥ 65) (Postel & Laurens, personal communication, May 2, 2019). Hence, a separate section for statements by elders was created (Appendix 5: Table 5; Table 6). In regard to positive statements, both elder participants indicated that conducting the AABMs was of little difficulty. Additionally, one elder participant stated that the game element is enjoyable for the age-group. Lastly, one elder participant indicated the sounds provided with correct or incorrect responses were supportive for elders.

Negative statements included that there are 'too many images that have to be selected for elders', and hence the research team should 'add multiple images of one drink for elders' to compensate for this. Specifically, it was stated that elders tend to know and enjoy significantly fewer alcoholic beverages than individuals of other age-groups. Hence, providing similar images of the same alcoholic drinks would not be hurtful to the enjoyment elders have in conducting the AABM, as one elder participant was 'unsure whether elders have an interest in personalisation' in general.

Discussion

Main findings

The purpose of the current study was to identify the convergent validity of the standardised and personalised AABMs included in *Breindebaas*. Specifically, the current study tempted to prove convergence between the two AABMs, as well as the convergence of the two AABMS individually with scores on other alcohol-related measurement instruments. Additionally, the current study analysed the relationship between neuroticism and alcohol-approach biases. In specific, the convergence between the standardised and personalised AABMs with a neuroticism-related measurement instrument was tested. Moreover, the current study addressed the possibility of moderation by neuroticism in the relationship between the obtained d-scores on the standardised and personalised AABMs. In sum, the current study did not find support for the five study questions included in the research, as the results on all study questions were found to be non-significant. Nevertheless, valuable information can be obtained from the current study that may support future AABM-research.

Study questions. Firstly, the current study was not able to obtain support for the convergence between the standardised and personalised AABMs (hypothesis 1). This finding suggests that the AABMs included in *Breindebaas* may measure different aspects of their shared underlying construct; the alcohol-approach bias. The finding was not expected, as the standardised and personalised AABMs are mobile translations of the R-AAT used by Wiers and others (2010[b]). However, after the exclusion of four participants who showed substantial differences in d-scores on the AABMs, significant findings were able to be obtained. Hence, there are indications that the standardised and personalised AABMs similarly measure the alcohol-approach bias. It is believed that the contrasting findings are a result of the lacking sample size. As the sample size of the current study shows to be relatively small, outliers may have had a large impact on the identified significance values.

Moreover, the current study could not establish convergence between d-scores on the individual AABMs, and scores on the Alcohol TLFB, AUDIT and the Craving VAS (hypothesis 2). Notably, this indicates that measures of alcohol-approach biases may not closely relate to measures of weekly amounts of alcohol consumed, levels of alcohol dependency and alcohol addiction, and the in-the-moment need to consume alcohol. The findings contrast those from the validity studies conducted by Kersbergen and others (2015) with the R-AAT. Specifically, these researchers found adequate levels of convergence between the R-AAT and the Alcohol TLFB and AUDIT. The difference in findings may be

explained by the sample size of the current study. Kersbergen and others (2015) included 80 participants in their validity research; the sample shows to be at least four times larger than the sample size of the current study. Additionally, the use of d-scores by the current study may have influenced the results. To illustrate, Kersbergen and others (2015) made use of both d-scores and bias scores based on raw RTs to identify whether the R-AAT could predict hazardous drinking and alcohol consumption. The use of d-scores led to insignificant results, as the R-AAT was found to be highly sensitive to methods that correct for outliers and errors in the data. Notably, the current study made use of a 2000ms cut-off to remove outliers and errors, as recommended by the AAT-guidelines of Nosek and others (2014). It may be possible that the cut-off point of 2000ms should be adjusted, either towards higher or lower values, as to establish statistical significance. It is believed that this optimal cut-off point has yet to be identified.

Furthermore, the current study could not confirm that the personalised AABM would provide more accurate (and thereby higher) d-scores than the standardised AABM (hypothesis 3). The hypothesis stemmed from the incentive-sensitisation theory of addiction (Robinson & Berridge, 2008), claiming that bias strengths are heavily dependent on the specific stimuli that initially led to their development. An indication for the non-significant finding is provided by the feedback points obtained from participants in the current study. In specific, the personalised AABM of *Breindebaas* requires one to select a minimum amount of fifty images in the ‘alcoholic’ and ‘non-alcoholic’ image categories. Nineteen participants indicated not to know or enjoy a minimum of fifty alcoholic beverages, and hence found the number of images to be selected too large. Of these 19 participants, nine participants stated to believe that the high number of images to-be-selected made the personalised AABM ‘unpersonal’. Thus, in light of the incentive-sensitisation theory of addiction, the personalised AABM may not cause strengthened bias levels as the stimuli shown are not fully relatable to the individual (Robinson & Berridge, 2008).

Lastly, the current study could not provide evidence for the convergence between levels of neuroticism and the strength of the alcohol-approach bias. This finding contradicts the tension reduction hypothesis, which states that highly neurotic individuals are more likely to consume more alcohol to reduce their levels of stress, and thereby develop stronger alcohol-approach biases (Liu et al., 2009). Additionally, the current study did not find evidence for the moderating effect of neuroticism on the association between d-scores on the standardised AABM and the personalised AABM, resulting from a greater need for

personalisation-options in highly neurotic individuals. The finding contrast those from Anagnostopoulou and others (2017), who identified that highly neurotic individuals may experience greater benefits of personalisation-options. A likely explanation for the two insignificant findings is that the current study found neuroticism scores to be unevenly distributed. Specifically, indications were provided that the sample scored higher or lower than expected on the BFI-N. The sample may thus be homogeneous in the neuroticism trait. Additionally, the sample shows to be relatively small to use for the conduct of statistical analyses. The lacking sample size may therefore have strengthened the effect of homogeneity on the significance of results. The insignificant difference found in d-scores on the AABMs as a result of moderation by neuroticism is believed to be caused by the lack of personalisation in the personalised AABM. Participants indicated that the personalised AABM included too many images, and hence deemed the personalised AABM to be unpersonal. Thus, it is likely that the personalised AABM was not beneficial enough for highly neurotic participants to take effect (Anagnostopoulou et al., 2017). Resultingly, this may have led highly neurotic participants to obtain equal d-scores as low-neurotic participants on the personalised AABM.

Feedback on Breindebaas by participants. The feedback obtained from participants was valuable in identifying possible future improvements for *Breindebaas*. As stated previously, concerns surrounding the number of images that had to be selected for the personalised AABM showed to be the most prominent remark within the sample. Nineteen of the 25 participants indicated a minimum-amount of fifty images to be too many. Interestingly, among the elder subgroup it was found that there may be no need for diversity in images. In contrast, elders indicated that the addition of images of the same drink under different contexts would be preferable. To the best of our knowledge, there exist no guidelines as of current regarding the minimum or maximum number of different images needed to find an effect. Therefore, the current study shall offer recommendations on the matter in the following sections.

Elders and d-scores. The current study encountered a controversial finding regarding the use of d-scores with the elder population. According to the AAT-guidelines proposed by Nosek and others (2014), participants should be excluded from further analyses when (more than) 25% of the responses are faster than 200ms, and slower than 2000ms. Resultingly, the current study had to exclude the two oldest participants of the sample. In addition, during previous research it was found that the R-AAT is sensitive to outlier correction methods, and as a result may provide insignificant findings (Kersbergen et al., 2015). The current study

included a sample of all age ranges, in which elders formed outliers due to providing slower-than-average responses. The findings indicate that elders may be consistently removed from the sample due to the AAT-guidelines, or to increase the chances of finding significant results. To the best of our knowledge, the aforementioned finding had not yet been addressed in previous research. Additionally, the current sample included solely two participants of elder age, and hence is believed to be too small to draw implications from. Therefore, it may be valuable to further address the use of d-scores with the elder population in future research, while making use of a larger sample including participants of all ages.

Strengths and limitations

The current study had several strengths as a result of its design and treatment of data. One such strength of the current study was the heterogeneity of the sample in its demographical characteristics. The previous study on the validity of the R-AAT was conducted with 80 students at a single university (Kersbergen et al., 2014). As a result, findings of the study may not be representative of the entire population, as the ages, educational levels, and possibly study directions of participants are homogeneous. The current study did make use of a convenience sampling procedure, leading participants to be selected on their availability. Nevertheless, an attempt was made to reach out to potential participants of different ages. Additionally, the current sample was found to exist of both students and working individuals of different educational levels. Resultingly, findings of the current study may be more applicable for generalization.

A further strength of the current study was the development of a quantified feedback table. Previous studies of the AABT in *Breindebaas* by Somsen (2017) included mentions of the feedback obtained as to improve the mobile application. Ultimately, this led to the inclusion of more diverse images of alcoholic and non-alcoholic drinks (Somsen, 2017). The current study tried to further improve the collection of feedback by labelling feedback points, and quantifying the number of participants who made comments that fitted the labels. This offered the current study the opportunity to identify numerous points of improvement. Additionally, by determining prominence of the feedback points based on the number of participants who shared the same view, it was more easily identified which adjustments should be made in a shorter period of time.

The current study also knew limitations. Firstly, the sample size of the current study showed to be limited in regard to the conduct of analyses. In specific, 27 participants were

initially included in the study, of which ten had to be excluded as a result of faulty data or exclusion criteria for the calculation of d-scores. Resultingly, the current sample knew solely 17 participants. This contrasts the validity studies on the R-AAT by Kersbergen and others (2015), who had a sample of 80 participants. In addition, the researchers controlled for faulty data by making use of estimation maximization algorithms (Kersbergen et al., 2015). As values were not missing at random in the current studies, but rather missing in its entirety, estimation maximization algorithms could not be used. The low sample size may have contributed to the insignificance of findings, as the analyses were found to be highly sensitive to outliers. To illustrate, values on the convergence between the standardised and personalised AABMs were found to change significantly by removing four outliers with disproportionate d-scores on both AABMs.

A second limitation of the study was the usage of the BFI-N to measure levels of neuroticism. The authors indicated not to develop norm tables for the BFI-N due to the inconsistency of personality trait levels over different samples. Contrastingly, it was recommended to establish means and standard deviations for the BFI-N based on the sample at hand (John & Srivastava, 1999). The current study has followed this advice, and established means and standard deviations based on the sample, as well as percentile norms for individual scores. As the current study's results were insignificant, the lack of norm tables of other samples complicated interpretations of the data. Comparisons could not be made as to whether the overall sample scored lower or higher in contrast to other samples. Resultingly, the current study lacks specific claims regarding the sample's extremity on the trait neuroticism, and its possible relationship with the observed d-scores.

Thirdly, the current study was limited by the technical difficulties experienced with *Breindebaas*. During the conduct of the AABMs, crashes occurred while swiping the images that caused the application to be unresponsive. The crashes may have increased the mean RTs of an unknown number of participants, and resultingly may have confounded the data. In addition, the crashes are also believed to have caused data to be missing in its entirety for some participants. The missing data ultimately led to a lacking sample size, which may explain part of the insignificance of the findings in the current study.

A fourth limitation of the current study were its inclusion criteria. The current study had the inclusion criteria: (1) participants are 18-years-old or above, and (2) male participants have to drink a minimum of seven drinks per week, and female participants have to drink a minimum of four drinks per week. Firstly, the minimum-requirements for weekly alcohol

consumption may have been too low to identify the convergent validity of the standardised and personalised AABMs with. This is, as the mean d-scores of the sample were found to near zero. Interestingly, the inclusion criteria used by Kersbergen and others (2015) knew lower minimum-requirements for alcohol consumption; participants were eligible to join if they consumed alcohol at least once per month. Nevertheless, the sample was found to consume 14 alcoholic drinks per week on average, whereas participants in the current study consumed 10 drinks. Secondly, the current study did not require participants to feel motivated to change their alcohol consumption levels. Motivation was found to be an important factor in reducing alcohol consumption levels with the AABT included in *Breindebaas* (Somsen, 2017). To the best of our knowledge, no indications exist whether motivation levels are implicated in the conduct of AABMs as well. Hence, it is uncertain whether a lack of motivation to change one's alcohol consumption levels may have had an effect on the obtained d-scores.

A last limitation of the current study was the degree of personalisation of the personalised AABM in *Breindebaas*. As mentioned before, a majority of the participants included in the current study believed that the minimum-requirement of fifty images to be selected in the alcoholic category was too many. Participants indicated not to enjoy, or not to recognise a minimum amount of fifty drinks. Resultingly, half of these participants experienced the AABM as unpersonal. The feedback by participants is believed to explain why the current study could not identify higher d-scores on the personalised AABM over the standardised AABM. The images may have failed to trigger one's alcohol-approach bias due to a lack of self-relevancy, as explained by the incentive-sensitisation theory of addiction (Robinson & Berridge, 2008). Additionally, the difficulty experienced in recognising images may have confounded the obtained d-scores, as this may have led to increases in mean RTs, as well as increased the variance within the sample.

Implications for practice and further research

The current study could not establish significance for the study questions included. Nevertheless, findings of the current study did prove valuable in multiple regards. Firstly, although the current study could not confirm the convergence between the standardised and personalised AABMs for the full sample, the exclusion of four outliers did lead to significant results. This provides indications that the insignificant findings are not a result of the inadequacy of AABMs in general. In contrast, it is likely that the insignificant findings were a result of other factors such as the current study's small sample size, the homogeneous

characteristics of the sample, and/or the difficulties experienced with *Breindebaas* (i.e. crashes; small images). The aforementioned factors may have likely affected the results for the other findings as well. Therefore, the current study encourages further improvement and use of the AABMs in *Breindebaas* in future studies, while emphasising on a need for larger sample sizes, to establish significance on the psychometric properties of the AABMs.

To establish significance, feedback provided by the current study's participants may be valuable to take into consideration. In specific, it is believed that an emphasis should be placed on feedback regarding the minimum and maximum-amount of diversity in images needed for the personalised AABM to be deemed 'personal'. Specifically, the current study found that the majority of participants did not know or enjoy a minimum amount of fifty alcoholic drinks. This finding may be connected to the insignificant results found in regard to the personalised AABM leading to higher d-scores. Indications were provided that participants may prefer images of the same drink in different positions or contexts. The indications were most prominent under the elder subgroup, in whom the belief existed that elders would not have an interest in diversity of images overall. Hence, it is believed that a second psychometric study of *Breindebaas* should be conducted in which there exists no minimum-number of drinks that have to be selected per category. A lack of personalisation can be compensated for by adding at least five images of the same drink in different positions or contexts.

In addition, questions arose in the current study on the inclusion criteria that were used. Firstly, the current study included male participants who consumed a minimum of seven alcoholic drinks per week, and female participants who consumed a minimum of four alcoholic drinks per week. The criteria were based on halving the norms for excessive alcohol consumption (CDC, n.D.). It is believed that this requirement may have been too low to establish the convergent validity of the standardised and personalised AABMs with. This is, as mean d-scores on both AABMs neared zero. Therefore, it may be fruitful for future studies on the AABMs in *Breindebaas* to shift the inclusion criteria regarding weekly alcohol consumption to 'excessive drinkers' (CDC, n.D.). In other words, male participants would have to drink at least fifteen alcoholic beverages per week, and female participants would have to drink at least seven alcoholic beverages per week, to be eligible to join the study.

Secondly, questions arose surrounding the importance of motivation to change one's alcohol consumption levels in the conduct of AABMs. It is believed that previous research was not conducted regarding the matter. Hence, adding the motivation to change one's

consumption levels as one of the inclusion criteria for future studies on the AABMs in *Breindebaas* is expected to be unsupportive as of current. Alternatively, a measure on the motivation to change one's alcohol consumption levels could be included in a questionnaire, and correlated with d-scores and other alcohol-related measures to establish its importance. If the measure is found to relate significantly to the obtained d-scores or other alcohol-related measures, further steps may be taken.

Lastly, the current study raised the question whether the use of d-scores with elders may be lacking in the conduct of AATs in general. The current study found elders to be either excluded due to AAT-guidelines, or due to the subgroup being more likely to form outliers in the sample. Notably, the current study solely included two elder participants. Hence, the size of the sub-sample shows to be too small to draw implications from. Therefore, it is believed that further research is needed, in which the adequacy of using d-scores in AATs with the elder subgroup is analysed based on a larger sample of different ages.

Conclusion

The current study was the first to assess the psychometric properties of the AABMs included in the mobile application *Breindebaas*, if not mobile AABMs in general. Furthermore, the current study was the first to identify the relationship between alcohol-approach biases and the need for personalisation in neurotic individuals. Although the main findings included in the study were found to be non-significant, important implications were identified. The current study contributed to the question surrounding the optimal balance between the diversity in images, and the degree of personalisation experienced by the user. Specifically, including images of the same drink in different positions, or with different backgrounds, may prove valuable in maintaining this balance. Further research is needed to explore whether the inclusion of multiple images of the same drink improves the experience of personalisation for users, and whether the improvement of the users' experience contributes to a change in the obtained d-scores.

References

Adang, R.P, Wensing, J. W. M, & Stockbrügger, R. W. (1998). Alcohol Consumption and Alcohol-related Liver Disease in The Netherlands. *Scandinavian Journal of Gastroenterology, 33*(225), 70-74. doi:10.1080/003655298750027263

Anagnostopoulou, E., Magoutas, B., Bothos, E., Schrammel, J., Orji, R., & Mentzas, G. (2017). Exploring the Links Between Persuasion, Personality and Mobility Types in Personalised Mobility Applications. *Persuasive Technology: Development and Implementation of Personalised Technologies to Change Attitudes and Behaviors*, 107–118. doi:10.1007/978-3-319-55134-0_9

Australian Government, Department of Veterans' Affairs (n.D). *Alcohol Screen (AUDIT)*. Retrieved May 12, 2019 from: http://nceta.flinders.edu.au/files/3314/2257/4957/Right_Mix_3.pdf

Babor, T.F., Higgins-Biddle, J.C., Saunders, J.B. & Monteiro, M.G. (2001). *AUDIT: The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Care (2nd edition)* [Measurement instrument]. World Health Organization, Geneva, Switzerland. Retrieved May 10, 2019, from: http://www.who.int/iris/handle/10665/67205

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182. doi:10.1037/0022 3514.51.6.1173

Boffo, M., Pronk, T., Wiers, R. W., & Mannarini, S. (2015). Combining cognitive bias modification training with motivational support in alcohol dependent outpatients: Study protocol for a randomised controlled trial. *Trials, 16*(1). doi:10.1186/s13063 015-0576-6

Carbia, C., Corral, M., Doallo, S., & Caamaño-Isorna, F. (2018). The dual-process model in young adults with a consistent binge drinking trajectory into adulthood. *Drug and Alcohol Dependence, 186*, 113-119. doi:10.1016/j.drugalcdep.2018.01.023

Centers for Disease Control and Prevention (n.D.). *What is excessive alcohol use?* Retrieved, March 29, 2019, from: https://www.cdc.gov/alcohol/pdfs/excessive_alcohol_use.pdf

Denissen, J. J., Geenen, R., Van Aken, M. A., Gosling, S. D., & Potter, J. (2008). Development and validation of a Dutch translation of the Big Five Inventory (BFI). *Journal of personality assessment, 90*(2), 152-157. doi:10.1080/00223890701845229

Eberl, C., Wiers, R. W., Pawelczack, S., Rinck, M., Becker, E. S., & Lindenmeyer, J. (2013).

Approach bias modification in alcohol dependence: Do clinical effects replicate and for whom does it work best? *Developmental Cognitive Neuroscience, 4,* 38-51. doi:10.1016/j.dcn.2012.11.002

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4,* 26–42. doi:10.1037/1040-3590.4.1.26

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26-34. doi:10.1037/0003-066X.48.1.26

Greenwald, A. G, Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197-216. doi:10.1037/0022-3514.85.2.197

Heitmann, J., Hemel-Ruiter, M. E., Vermeulen, K. M., Ostafin, B. D., Macleod, C., Wiers, R. W., & Jong, P. J. (2017). Internet-based attentional bias modification training as add on to regular treatment in alcohol and cannabis dependent outpatients: A study protocol of a randomized control trial. *BMC Psychiatry, 17*(1). doi:10.1186/s12888 017-1359-2

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives [Measurement instrument]. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 102–138). New York: Guilford Press.

Kersbergen, I., Woud, M. L., & Field, M. (2015). The validity of different measures of automatic alcohol action tendencies. *Psychology of Addictive Behaviors, 29*(1), 225 230. doi:10.1037/adb0000009

Kim, S. S., Gulick, E. E., Nam, K. A., & Kim, S.-H. (2008). Psychometric Properties of the Alcohol Use Disorders Identification Test: A Korean Version. *Archives of Psychiatric Nursing, 22*(4), 190–199. doi:10.1016/j.apnu.2007.07.005

Lac, A., & Donaldson, C. D. (2016). Alcohol attitudes, motives, norms, and personality traits longitudinally classify nondrinkers, moderate drinkers, and binge drinkers using discriminant function analysis. *Addictive Behaviors, 61,* 91–98. doi:10.1016/j.addbeh.2016.05.006

Leung, D. Y., Wong, E. M., Chan, S. S., & Lam, T. (2012). Psychometric properties of the Big Five Inventory in a Chinese sample of smokers receiving cessation treatment: A validation study. *Journal of Nursing Education and Practice, 3*(6). https://doi.org/10.5430/jnep.v3n6p1

Liu, S., Wang, M., Zhan, Y., & Shi, J. (2009). Daily work stress and alcohol use: testing the

cross-level moderation effects of neuroticism and job involvement. *Personnel Psychology, 62*(3), 575–597. doi:10.1111/j.1744-6570.2009.01149.x

Maasland, D. H., Brandt, P. A., Kremer, B., Goldbohm, R. A., & Schouten, L. J. (2014). Alcohol consumption, cigarette smoking and the risk of subtypes of head-neck cancer: Results from the Netherlands Cohort Study. *BMC Cancer, 14*(1). doi:10.1186/1471 2407-14-187

Manning, V., Staiger, P. K., Hall, K., Garfield, J. B., Flaks, G., Leung, D.,Verdejo, & Garcia, A. (2016). Cognitive Bias Modification Training During Inpatient Alcohol Detoxification Reduces Early Relapse: A Randomized Controlled Trial. *Alcoholism: Clinical and Experimental Research, 40*(9), 2011-2019. doi:10.1111/acer.13163

Meneses-Gaya, C., Zuardi, A. W., Loureiro, S. R., & Crippa, J. A. S. (2009). Alcohol Use Disorders Identification Test (AUDIT): An updated systematic review of psychometric properties. *Psychology & Neuroscience, 2*(1), 83–97. doi:10.3922/j.psns.2009.1.12

Moss, A. C., & Albery, I. P. (2009). A dual-process model of the alcohol–behavior link for social drinking. *Psychological Bulletin, 135*(4), 516-530. doi:10.1037/a0015991

Nijen Es, L. (2017). Alcohol Cognitive Bias Modification training voor probleemdrinkers via de smartphone: een pilot studie follow-up. *Master dissertation, August 2017, University of Twente, Enschede*

Nosek, B. A., Bar-Anan, Y., Sriram, N., Axt, J., & Greenwald, A. G. (2014). Understanding and Using the Brief Implicit Association Test: Recommended Scoring Procedures. *PLoS ONE, 9*(12). doi:10.1371/journal.pone.0110938

Owusu, D., Quinn, M., & Wang, K. S. (2015). Alcohol Consumption, Depression, Insomnia and Colorectal Cancer Screening: Racial Differences. *International Journal of High Risk Behaviors and Addiction, 4*(2). doi:10.5812/ijhrba.4(2)2015.23424

Papachristou, H., Nederkoorn, C., & Jansen, A. (2016). Neuroticism and Negative Urgency in Problematic Alcohol Use: A Pilot Study. *Substance Use & Misuse, 51*(11), 1529-1533. doi:10.1080/10826084.2016.1178294

Qualtrics (2018) [Computer software]. Provo, UT. Retrieved May 10, 2019, from: https://www.qualtrics.com/

Random Lists (n.D.). *Random Team Generator*. Retrieved March 20, 2019, from: https://www.randomlists.com/team-generator

Robinson, T. E., & Berridge, K. C. (2008). The incentive sensitization theory of addiction: some current issues. Philosophical Transactions of the Royal Society B: *Biological*

Sciences, 363(1507), 3137–3146. doi:10.1098/rstb.2008.0093

Salemink, E., Van Lier, P., Meeus, W., Raaijmakers, S., & Wiers, R. (2015). Implicit alcohol relaxation associations in frequently drinking adolescents with high levels of neuroticism. *Addictive Behaviors, 45*, 8–13. doi:10.1016/j.addbeh.2015.01.002

Schippers, G.M., & Broekman, T.G. (2010). *AUDIT: Alcohol Use Disorders Identification Test. Nederlandse Vertaling*. Retrieved May 11, 2019, from: https://www.mateinfo.nl/audit/audit-nl.pdf

Schwandt, M. L., Heilig, M., Hommer, D. W., George, D. T., & Ramchandani, V. A. (2012). Childhood Trauma Exposure and Alcohol Dependence Severity in Adulthood: Mediation by Emotional Abuse Severity and Neuroticism. *Alcoholism: Clinical and Experimental Research, 37*(6), 984–992. doi:10.1111/acer.12053

Sobell, L. C., Brown, J., Leo, G. I., & Sobell, M. B. (1996). The reliability of the Alcohol Timeline Followback when administered by telephone and by computer. *Drug and Alcohol Dependence, 42*(1), 49–54. doi:10.1016/0376-8716(96)01263-X

Sobell, L. C., & Sobell, M. B. (1992). Timeline Follow-Back [Measurement instrument]. *Measuring Alcohol Consumption*, 41–72. doi:10.1007/978-1-4612-0357-5_3

Sobell, L.C. & Sobell, M.B. (1995). *Alcohol Timeline Followback Users' Manual*. Toronto, Canada: Addiction Research Foundation.

Somsen, B.N. (2017). Alcohol Cognitive Bias Modification training for problem drinkers via their smartphone: a pilot study. *Master dissertation, March 2017, University of Twente, Enschede*

Srivastava, S. (2012). *Norms for the Big Five Inventory and other personality measures*. Retrieved May 14, 2019, from https://thehardestscience.com/2012/10/17/norms-for-the-big-five-inventory-and-other-personality-measures/

Tactus Verslavingszorg (n.D.). *Handige apps*. Retrieved, February 26, 2019, from: https://www.tactus.nl/Advies-en-Hulp/Handige-apps

Wiers, R. W., Ames, S. L., Hofmann, W., Krank, M., & Stacy, A. W. (2010[a]). Impulsivity, Impulsive and Reflective Processes and the Development of Alcohol Use and Misuse in Adolescents and Young Adults. *Frontiers in Psychology, 1*. doi:10.3389/fpsyg.2010.00144

Wiers, R. W., Rinck, M., Kordts, R., Houben, K., & Strack, F. (2010[b]). Retraining automatic action-tendencies to approach alcohol in hazardous drinkers. *Addiction, 105*(2), 279 287. doi:10.1111/j.1360-0443.2009.02775.x

Wiers, R. W., Houben, K., Fadardi, J. S., Beek, P. V., Rhemtulla, M., & Cox, W. M. (2015).

Alcohol Cognitive Bias Modification training for problem drinkers over the web. *Addictive Behaviors, 40*, 21-26. doi:10.1016/j.addbeh.2014.08.010

World Health Organization (WHO) (2014). *Netherlands*. Retrieved February 15, 2019, from: https://www.who.int/substance_abuse/publications/global_alcohol_report/profiles/nld pdf

World Health Organization (WHO) (2018). *Global status report on alcohol and health 2018*. Retrieved February 16, 2019, from: https://apps.who.int/iris/bitstream/handle/10665/274603/9789241565639 eng.pdf?ua=1

Zeegers, M. P., Volovics, A., Dorant, E., Goldbohm, R. A., & Brandt, P. A. (2001). Alcohol Consumption and Bladder Cancer Risk: Results from the Netherlands Cohort Study. *American Journal of Epidemiology, 153*(1), 38-41. doi:10.1093/aje/153.1.38

Appendix

Appendix 1: Flyer Breindebaas-study

Hou jij ook van een borrel?

En zou jij de wetenschap een handje willen helpen?
Dan zoeken wij jou!

WIE: mannen (18+) die minimaal 7 glazen alcohol per week drinken en vrouwen (18+) die minimaal 4 glazen alcohol per week drinken

WAT: ter plekke uitvoeren van 2 verschillende testen in de 'Breindebaas' app (totaal duur: ± 30 min.)

WAAR: Universiteit Twente (andere locatie is in overleg mogelijk)

WANNEER: 16 april - 10 mei 2019

BELONING: onder iedere 10 deelnemers wordt een 'bol.com' cadeaubon ter waarde van €20 verloot

INTERESSE: mail naar utjulietteniels@gmail.com

Appendix 2: Standardised texts research procedure

1. Pre-measure:

1.1. Questionnaire:

'Bedankt voor [je/uw] deelname aan dit onderzoek. Allereerst gaan we straks beginnen met de vragenlijst. Aan het begin van de vragenlijst staat een toestemmingsformulier; hierin worden zaken uitgelegd over het gebruik van de data, privacy, en de behandeling van opmerkingen of klachten. Wij vragen [je/u] deze zorgvuldig door te lezen voordat er toestemming gegeven wordt. Mochten [je/u] nog verdere opmerkingen of vragen hebben (betreffende het toestemmingsformulier), beantwoorden wij ze graag.'

English translation. 'Thank you for your participation in our study. Firstly, we will start with the questionnaire. At the start of the questionnaire, the consent form can be found; herein things will be explained about the use of the data, privacy, and the treatment of comments or complaints. We ask of you to carefully read this before consent is given. May you have further comments or questions regarding this (consent form), feel free to state them.'

1.2. Logging on:

'Nu gaan we bezig met de app. Voordat we beginnen, willen we [je/u] graag vragen om het geluid van de telefoon aan te zetten op 50%.'

'Dan kunnen we nu beginnen met het aanmaken van een account. Om een account aan te maken, mag [je/u] klikken op de tekst onderaan het (start)scherm, en de gevraagde informatie invullen.'

'[Je/u] mag [zich/je] nu aanmelden.'

English translation. 'Now we will make use of the app. Before we start, we would like to ask you to turn the phone's sound on to 50%.'

'Now you may start with the creation of an account. To create an account, you may click on the text shown at the bottom of the (start)screen, and fill in the requested information.'

'You may now log on.'

2. Measures:

2.1. Personalised trial:

‘Voor de start van deze ronde, vragen we [je/u] de drankjes te selecteren die u mogelijk wel zou drinken uit de verschillende categorieën. We raden [je/u] aan om, wanneer [je/u] een afbeelding selecteert, een andere te verwijderen. Dit komt omdat de app een minimum- en maximumaantal afbeeldingen kent.’

‘Dan gaan wij nu beginnen met de meting. De app zal voor het begin van de meting aangegeven wat er van [je/u] verwacht wordt. Veel succes!’

English translation: ‘Before the start of this round, we would like you to select the drinks you would be likely to drink from the different categories. We would like to recommend you to, might you select an image, to remove another one in turn. This is, as the app knows a minimum- and maximum number of images needed.’

‘Now we can start with the measurement. De app will tell you at the start of the measurement what is expected of you. Good luck!’

2.2. Standardised trial:

‘Gedurende deze ronde gebruiken we de standaard afbeeldingen. Er hoeven dus geen afbeeldingen gekozen te worden. De app geeft voor de meting aan wat van [je/u] verwacht wordt. Succes!’.

English translation: ‘During this round we will use the standard images. This means that you will not have to select pictures yourself. Before the measurement, the app will tell you what is expected of you. Good luck!’

2.3. Break (after the first measurement):

‘De eerste meting is nu voltooid. Voordat we met de volgende meting beginnen, zullen we tien minuten met iets anders bezig gaan. [Je/u] hebt de mogelijkheid om in deze tien minuten met een puzzel aan de slag te gaan. Als [je/u] liever iets anders doet, mag [je/u] ook met [je/u] mobiel bezig, of kunnen we een gesprekje tussendoor hebben.’

English translation: ‘The first measurement is now finished. Before we start with the next one, we will have a break of ten minutes. We have brought with a puzzle for you to finish, if you would like to do so. If you prefer to do something else, you can also check your phone or have a conversation with us.’

3. Post-measure:

‘We zijn nu aangekomen aan het einde van het onderzoek. We willen [je/u] nogmaals hartelijk bedanken voor het deelnemen aan het onderzoek. Kort willen we [je/u] nog even vragen hoe [je/u] het onderzoek [hebt/heeft] ervaren.’

English translation: ‘We have now reached the end of the research. We would like to thank you, again, for your participation in the research. To end, we would like to ask you how you have experienced the research.’

Appendix 3: D-score calculations for the standardised and personalised AABM

Appendix Table 1

Calculation D-scores Standardised AABM

Part. nr.	Alcohol AV* Mean	Alcohol AP** Mean	Alcohol AV SD	Alcohol AP SD	Pooled Alcohol SD	D-scores
1	1111.35	1079.32	121.43	77.03	101.68	.32
2	1063.45	1053.67	70.85	86.58	79.11	.12
3	1102.55	1136.78	114.30	130.01	122.41	-.28
4	1205.89	1258.41	214.09	152.70	185.94	-.28
5	1728.44	1342.19	362.96	236.50	306.32	1.26
6	1294.82	1251.42	176.95	169.65	173.34	.25
7	1152.12	1162.94	174.03	144.87	160.12	-.07
8	1136.85	1251.22	194.68	345.67	280.53	.31
9	1170.75	1175.57	154.24	88.14	125.62	-.04
10	1208.18	1205.70	185.58	220.91	204.01	.01
11	1534.18	1371.47	249.89	153.82	207.49	.78
12	1147.80	1265.63	169.82	103.38	140.59	-.84
13	1262.18	1226.70	281.57	220.91	253.07	.14
14	1175.26	1148.72	179.52	234.53	208.84	.13
15	1206.06	1206.27	148.07	205.60	179.16	.00
16	1337.64	1271.82	175.65	230.37	204.84	.32
17	1172.90	965.90	231.49	77.78	172.68	1.20

Note. Alcohol AP = Alcohol-Approach. Alcohol AV = Alcohol-Avoidance. The Alcohol D-scores have been calculated with the formula: $(M_{RT\text{-}push\text{-}alcohol} - M_{RT\text{-}pull\text{-}alcohol})/SD_{alcohol}$.

Appendix Table 2

Calculation D-scores Personalised AABM

Part. Nr.	Alcohol AV Mean	Alcohol AP Mean	Alcohol AV SD	Alcohol AP SD	Pooled Alcohol SD	D-scores
1	1074.42	1182.80	80.72	182.83	141.32	-.77
2	1136.05	1028.19	185.60	94.15	147.16	.73
3	1100.35	1091.79	120.02	240.19	189.86	.05
4	1143.21	1270.50	149.03	230.12	193.86	-.66
5	1130.36	1145.13	172.45	189.32	181.08	-.08
6	1155.65	1158.15	161.80	162.68	162.24	-.02
7	1048.24	1066.74	84.40	95.19	89.96	-.21
8	1408.79	1389.87	243.45	258.01	250.84	.08
9	1185.68	1175.65	76.96	51.97	65.67	.15
10	1356.91	1233.05	285.04	157.98	230.44	.54
11	1439.16	1341.30	259.27	188.39	226.62	.43
12	1149.46	1071.14	106.82	106.96	106.89	.73
13	1206.50	1162.11	216.21	217.60	216.91	.20
14	1367.15	1095.22	311.28	135.97	240.19	1.13
15	1309.34	1284.10	258.68	178.96	222.42	.11
16	1317.91	1331.00	252.04	218.72	235.97	-.06
17	1130.30	1012.67	152.28	197.59	176.39	.67

Note. Alcohol AP = Alcohol-Approach. Alcohol AV = Alcohol-Avoidance. The Alcohol D-scores have been calculated with the formula: $(M_{RT\text{-}push\text{-}alcohol} - M_{RT\text{-}pull\text{-}alcohol})/SD_{alcohol}$.

Appendix 4: Participants' individual scores on the measurement instruments

Appendix Table 3

Participants' Individual Scores on the Alcohol-related and Neuroticism-related Measurement Instruments

Part. nr.	Alcohol TLFB			AUDIT			Craving VAS	BFI-N		
	Score	Percentile	Label	Score	Percentile	Risk level		Score	Percentile	Label
1	8.00	32.35	Low	8.00	55.88	Moderate	6.00	29.00	79.41	High
2	10.00	55.88	Low	6.00	35.29	Low	4.00	26.00	64.71	High
3	14.00	76.47	High	8.00	55.88	Moderate	7.00	32.00	91.18	High
4	7.00	14.71	Low	4.00	5.88	Low	.00	33.00	100.00	High
5	8.00	32.35	Low	8.00	55.88	Moderate	.00	32.00	91.18	High
6	3.00	5.88	Low	12.00	85.29	Moderate	.00	23.00	29.41	Low
7	17.00	91.18	High	11.00	76.47	High	.00	22.00	17.65	Low
8	8.00	32.35	Low	5.00	20.59	Low	5.00	23.00	29.41	Low
9	17.00	91.18	High	8.00	55.88	Moderate	7.00	17.00	5.88	Low
10	21.00	100.00	High	15.00	94.12	Moderate	.00	25.00	55.88	High
11	10.00	55.88	Low	7.00	41.18	Low	.00	27.00	70.59	High
12	15.00	82.35	High	12.00	85.29	Moderate	1.00	24.00	44.12	Low
13	9.00	47.06	Low	5.00	20.59	Low	.00	8.00	11.76	Low
14	7.00	14.71	Low	5.00	20.59	Low	1.00	25.00	55.88	High
15	11.00	64.71	High	5.00	20.59	Low	6.00	23.00	29.41	Low
16	8.00	32.35	Low	16.00	100.00	High	.00	29.00	79.41	High
17	12.00	70.59	High	9.00	70.59	Moderate	2.00	24.00	44.12	Low

Note. Labels on the Alcohol TLFB and BFI-N have been established by identifying scores above the sample mean as 'High', and below the sample mean as 'Low'. Risk levels on the AUDIT were established by using the norm tables provided by the Australian Government, Department of Veterans' Affairs (n.D.). The Alcohol TLFB is developed by Sobell and Sobell (1995). The AUDIT is developed by Babor and others (2001). The BFI-N is developed by John and Srivastava (1999).

Appendix 5: Quantified feedback table

Appendix Table 4

Feedback Points obtained by Participants on the Breindebaas-app (*n* = 25)

Participant no.	Condition	Comments
P1	P	- Enjoyable game-element (+) [1] - Doing the measure creates thirst for alcohol (-) [11]
P2	P	- Not sure whether there were tricky images, such as alcohol-free beer (N) [19]
P3	P	- Too many choices (-) [2] - Use texts instead of images in-measure (N) [3] - Doing the measure creates thirst for alcohol (=) [11]
P4	P	- Too many choices (-) [2] - Add multiple images of one drink (-) [13]
P5	P	- Enjoyable game-element (+) [1] - Too many choices (-) [2] - Categorise images per brand/type of drink (-) [5]
P6	P	- Too many choices (-) [2] - Categorise images via text (-) [5] - Select button of images unresponsive (-) [6]
P7	P	- Too many choices/not personalised (-) [7] - Categorise images per brand/type of drink (-) [5] - Images are double-up (-) [12] - Not sure whether there were tricky images, such as alcohol-free beer (N) [19]
P8	P	- Too many choices (-) [2] - Button of images unresponsive (-) [6] - Categorise images per brand/type of drink (-) [5] - Not sure whether there were tricky images, such as alcohol-free beer (N) [19]
P9	P	None
P10	P	- Doing the measure creates thirst for alcohol (-) [11]
P11	S	- Too many choices/not personalised (-) [7] - Small images/increase size (-) [4] - Images are double-up (-) [12]
P12	S	None
P13	S	- Add Sambuca (N) [9]
P14	S	- Too many choices (-) [2] - Categorise images per brand/type of drink (-) [5] - Add multiple images of one drink (-) [13]
P15	S	- Too many choices (-) [2] - Categorise images per brand/type of drink (-) [5] - Use texts instead of images in-measure (N) [3] - Remove background during measure (-) [10]
P16	P	- Too many choices/not personalised (-) [7]

		- Doing the measure creates thirst for alcohol (-) [11] - Images are double-up (-) [12] - Not sure whether there were tricky images, such as alcohol-free beer (N) [19]
P17	S	None
P18	S	- Too many choices/not personalised (-) [7] - Not all images were clear (-) [20] - The app said before a measurement that you had to swipe 'soft drinks', but what to do with drinks like coffee and tea (-) [21]
P19	S	- Not sure whether there were tricky images, such as alcohol-free beer (N) [19] - Too many choices/not personalised (-) [7]
P20	P	- Too many choices (-) [2] - Too many choices/not personalised (-) [7] - Images are double-up (-) [12] - Not all images were clear (-) [20]
P21	P	- Enjoyable game-element (+) [1] - Not sure whether there were tricky images, such as alcohol-free beer (N) [19] - Too many choices/not personalised (-) [7] - The app said before a measurement that you had to swipe 'soft drinks', but what to do with drinks like coffee and tea (-) [21]
P22	P	- Not sure whether there were tricky images, such as alcohol-free beer (N) [19] - Too many choices/not personalised (-) [7] - Images are double-up (-) [12] - Not all images were clear (-) [20]
P23	P	- Too many choices (-) [2] - Too many choices/not personalised (-) [7] - Images are double-up (-) [12]

Appendix Table 5

Feedback Points obtained by Elder Participants on the Breindebaas-app (*n* = 2)

P24	S	- Measure is easy to do for elders (+) [13] - Add multiple images of one drink for elders (-) [14] - Too many images that have to be selected for elders (-) [15] - Unsure whether elders have an interest in personalisation (-) [16] - Sounds support memory for elders (+) [17]
P2	S	- Measure is easy to do for elders (+) [13] - The game element is enjoyable for elders (+) [21]

Appendix Table 6

Labeled Feedback Points obtained by Participants on the Breindebaas-app (*N* = 27)

Labels	Feedback points
Positive (+)	[1] Enjoyable game element (3 p.)
Neutral (N)	[3] Use texts instead of images in-measure (2 p.) [8] Add Sambuca (1 p.) [19] Uncertainty whether there were tricky images, such as alcohol-free beer (7 p.)
Negative (-)	[2] Too many images to select (9 p.) [7] Too many images to select/not personalised (10 p.) [4] Small images/increase size (3 p.) [6] Select button of images unresponsive (2 p.) [5] Categorise images per brand/type of drink (5 p.) [9] Remove background during measure (1 p.) [11] Images are double-up (6 p.) [12] Add multiple images of one drink (3 p.) [19] Not all images were clear (3 p.) [20] The app said before a measurement that you had to swipe ‘soft drinks’, but what to do with drinks like coffee and tea (2 p.) [10] Doing the measure creates thirst for alcohol (4 p.)
Elders (E)	[13] Measure is easy to do for elders (2 p.) [14] Add multiple images of one drink for elders (1 p.) [15] Too many images that have to be selected for elders (1 p.) [16] Unsure whether elders have an interest in personalisation (1 p.) [17] Sounds support memory for elders (1 p.) [21] The game element is enjoyable for elders (1 p.)