Visual Perception of international traffic signs: Influence of teaching materials and culture on eye movements

GERGELY RAKOCZI, Vienna University of Technology, Institue for Design and Assessment of Technology, Vienna, Austria
ANDREW T. DUCHOWSKI, Clemson University, School of Computing, Clemson, SC, USA
HELENA CASAS-TOST, Universitat Autònoma de Barcelona, Traducció i Interpretació, Barcelona, Spain
MARGIT POHL, Vienna University of Technology, Institue for Design and Assessment of Technology, Vienna, Austria

General Terms: Visual Perception, Signage, Culture

Additional Key Words and Phrases: Eye Tracking, e-Learning, Visual Behaviour, Learning Instructions

1. INTRODUCTION

Driver education consists of two major parts, the theoretical and the practical training. The former includes the learning of traffic signs, which is commonly assissted by teaching programs. For exercising and practising of traffic signs commonly e-learning environments is applied. The subject of this paper introduces an experiment that focuses on the visual perception of traffic signs that are embedded into e-learning systems.

Eye tracking research on the context of driving is almost exclusively concentrating on the investigation of practical driving. There is much research on eye tracking devices for the detection of driver's physical condition like fatigue ([Zhu and Ji 2004]), supporting drivers at recognition of traffic signs in different driving environments (as reference see [Ko et al. 2010]), investigating conspicuity and capability of recalling traffic signs during driving ([Inman 2012]) or tracking drivers' eyes in driving simulators, such as [Fisher et al. 2006], who reported an eye tracking study including a driving simulator for risk awareness on novice drivers. Others focus on measuring driver's cognitive load during driving performance in order to be able to remotely detect driver's distraction ([Palinko et al. 2010]). However to sum it up, there is hardly any research on eye movements of the visual perception of traffic signs during a learning/practising scenario, which is also a crucial part of driving. This paper tries to close this gap.

Related work in general on the perception of traffic signs is done by [Ng and Chan 2007]. Within a questionnaire study [Ng and Chan 2007] examined five cognitive design aspects of Chinese traffic signs such as familiarity, concreteness, simplicity, meanigfulness and semantic closeness. Findings show that frequency of encounter of the road signs' icons significantly influence semantic closeness. Simply put, the more people see specific traffic signs the more they know their meanings. For this the study reveals evidency that the sign's meaningfulness have strong statistical correlation with sign familiarity. This study also yields that concrete sign icons—such as direct visualisations of objects—help people to elicit meaning and to link it to the traffic sign's function more easily. However, [Ng and Chan

Contact author's address: A. T. Duchowski, Clemson University, School of Computing, 100 McAdams Hall, Clemson, SC 29634, USA.
This work was partly supported by a grant of Vienna University of Technology for short-term academic work abroad.

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

2007]'s work did not investigate coloring, shape or texts displayed on signs, nor did the study collect any eye data of the participants.

An other related work to perception and traffic signs is done by [Yan et al.]. This study aims to predict the visual perception of traffic signs based on specific characteristics. A model consisting of visual attentional model, Gaussian pyramid creation, visual feature extraction and saliency map are used to generate driver's eye fixations. The findings of the study show good prediction results for both letters and symbols depicted in traffic signs, however is not suffienct enough to cover complex traffic signs.

Beside the investigation of traffic signs there is a second focus of this paper. Within the presented study the influence of teaching lessons prior online tests is evaluated. There are numerous reports on the effect that various task instructions provoke different eye movement. As an example see [Kaakinen and Hyönä 2010] where task instructions were varied in order to see its effects on reading. Also [Holmqvist et al. 2011] lists many eye tracking experiments were task instructions caused significant differences in the visual perception of entitites. [Taya et al. 2012] did research on the investigation of different goal during percieving complex dynamic natural scenes, such as clip of sport events. However, there are no investigations in the field of e-learning. In this paper a specific design is considered in order to be able to determine differences of eye movements that are caused by different level of preparation of learners prior online tests.

At last also cultural aspects are considered in test design of this paper, as aiming to investigate cross-cultural differences. Within this experiment eye movements of three ethnic groups from three different continents are considered. There are participants involved from Europe (Austria), America (USA) and Asia (China). Related work in between cultural differences and eye movements are numerous in the field of marketing. As an example for a large-scale investigation see [Peeters 2010]. Finding of this study state that different ethnic groups have various information needs and also that specific cultures give more visual value to different parts of information. As the experiment was carried out on e-commerce websites the findings might not be valid for e-learning environments.

An other related work—conducted by [Chua et al. 2005]—reveal cultural differences in general scene perception. The key essence of this study yields that Easterns and Westerns elicit different visual pathes in terms of percieving fore- and background objects of certain scenes. Americans fixate faster and longer on foregrounded objects extracting hereby more visual details than Chinese, whereas Eastern participants tend to have eye movements more holistically over the screen. However, one limitation of this study is that the tasks were focusing on free scene exploration, hereby only marginally narrowing users' visual tasks. To ensure a realistic investigation scenario for each ethnic group, the e-learning course and teaching materials were presented in the language of the cultures. Therefore for implementing the Chinese verion of the e-learning course recommendations given by [Yen et al. 2011] were considered.

2. BACKGROUND

2.1 Eye Tracking

Eye tracking is applied as the main data gathering method of the investigation described in this paper. Eye tracking as a methodology aims to collect eye movements of participants looking at a computer screen or a specificially defined region in space. Eye movements themselves are defined as changes in the direction of participant's gaze during persuing objects of interest. The theoretical fundament for using eye tracking data is based on the assumption that there is a relationship between one's eye movements and the corresponding cognitive processes [Just and Carpenter 1976]. This correlation implies that those areas on the screen or regions in space are fixated, which attract the user's attention. Furthermore eye movements are also appropriate to link users' visual explorations to cognitive processes during problem or task solving. Also as [Bente 2005] states eye movements may be regarded as valuable indicators for specific cognitive conditions e.g. attention, interest, complexity, etc. that may occur during learning. The challange however, is to draw fairly valid conclusions from the eye tracking data on the user's thought processes. Eye tracking as an input in the research area of e-learning is not commoly used. For an overview of exemplary application scenarios of eye tracking in e-learning see [Rakoczi 2012]. In general there is little eye tracking research in e-learning and most of the existing investigations rather focus on the development of adaptive e-learning systems than on the investigation of learning processes. The study reported in this paper is a contribution in this area.

However, similar to other evaluation methodologies, also eye tracking is limited. The most fundamental and most discussed limitations of eye tracking are reported by [Jacob and Karn 2003]. Crutial limitation for the study described in this paper are as followed. Many eye tracking researchers argue that the "'spotlight of attention"' assumption is a significant problem for data analysis, as it is possible to fixate an areo of interest without cognitively processing it. Therefore gazeplots might reflect users' visual path, however it cannot ensure that the participant actually cognitively processes fixated items. [Bente 2005] further expanded this thought by stating that eye tracking data, also cannot provide sufficient explanation for non-fixations on certain areas. Therefore according to [Nielsen and Pernice 2010] data analysis and interpretation of eye movements in terms of concluding general statements needs high expertise and experience. As there are no universal standards for interpretation and only a few basic approaches for correlation frameworks, data analysis has to be conducted with distinctive carefulness [Holmqvist et al. 2011]. Also, the more complex an eye tracking design scenario gets the more "'noisy"' the results are. Especially in eye tracking, the probability that participants get cognitively distracted is higher the longer test sessions last. Therefore task design must take this in account, as reported by [Pernice and Nielsen 2009].

In the present study eye movements were recorded to observe visual perception of international traffic signs. General variables, such as task success rate and time completion are used to detect differences. Eye tracking variables are used in order to deduce further explaination on participants' visual perception. Definition, interpretation and hypothesis for each eye tracking metrics are described in the following sections 2.2. As reference [Holmqvist et al. 2011], [Ehmke and Wilson 2007], [Nielsen and Pernice 2010] and [Goldberg and Kotval 1999] were considered.

2.2 Eye tracking metrics

2.2.1 *Overall regression rate on traffic signs.* This metrics describes the absolute rate of participants' re-visits on traffic signs relative to number on traffic signs for each test item. Re-visits stand for the total count of fixations within the AOI that covers each traffic sign's image. Re-visits are calculated by taking the total fixation value minus 1, as the AIO has been already processed once. If traffic signs AOIs have a total fixation count of zero, the regression rate also remains zero, as no negative values are possible.

$$\left|\frac{\sum_{i=1}^{N} AOI_i - 1}{N}\right|$$

(N ... number of traffic signs within the test item)

Eye movements' re-visits on AOIs usually indicate a higher cognitive investigation in terms of information extraction. Often regression stands for difficulty of the visual object or reveals higher visual comparison, that identifies competing elements within the stimulus. Also, regressions might be elicited in order to refresh short-term working memory. For the present study it is expected that the overall regression rate is lower for domestic than for foreign traffic sign. Furthermore regression rate is expected to be higher in mixed task type scenarios and lower for members of the treatment group. There are no differences estimated among the three ethnic groups.

2.2.2 *Totel fixation count.* This eye tracking variable stands for the total number of valid fixations within an AOI, ignoring thereby the fixation durations. If at the end of the test session the participant has not fixated on the AOI, the fixation count value will be zero. In general total fixation count defines semantically informative. For task solving processes total fixation count indicates regions where higher cognitive activity occured. Simply put, total fixation count identifies areas of general importance. However, for search tasks total fixation count is negatively correlated as low values identify that the goal of task is easy to reach or that the participant in very experienced. Also, in the conext of search tasks high values indicate difficulty in interpreting the stimulus. On foreign signs for matching and true/false tasks would generate higher values, whereas on domestic signs lower fixation counts are expected. For search tasks the opposite effect is anticipated. The influence of teaching materials should provoke faster task solving, therefore lower fixation counts are predicted for the overall test sessions for members of the treatment group. No prediction can be given on the fixation counts of the ethnic groups.

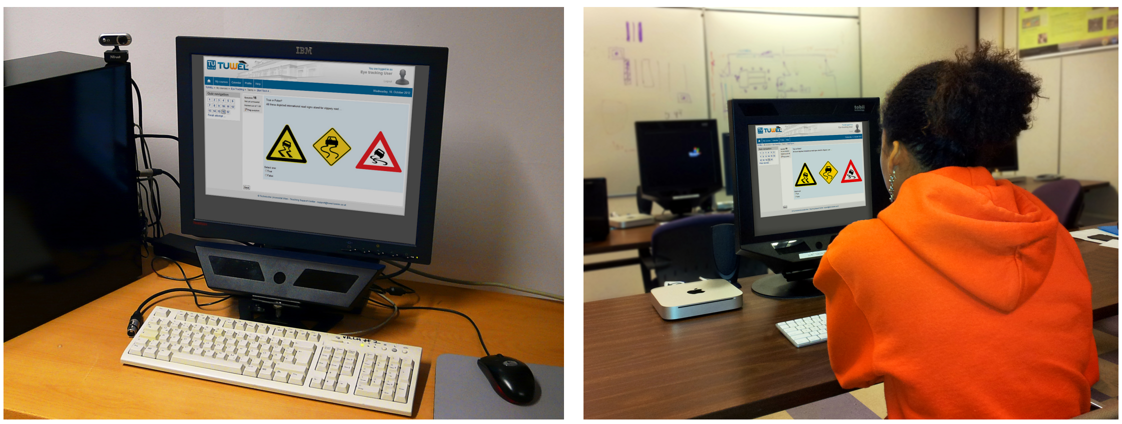
2.2.3 *Mean fixation duration.* This parameter describes the arithmetic average fixation durations of participicants on a specific AOI. According to the literature longer fixation durations equal to more effortful cognitive processing of visual stimulus. Also unexpected, such as out-of-context objects, less frequently encountered elements or unclear/blurred parts of images generate longer fixations. Also high mental efforts, such as calculations or driving on high incident curves or driving at night provoke longer mean fixation duration. In usability longer fixation durations indicate difficulty in information exctraction or interpretation of meaning from fixated objects on the screen can be deduced. However, reports show that non-demanding elements can push participants towards daydreaming, provoking hereby longer fixations that in this case stand for shallow information processing or low levels of concentration. Also high stress, impatience or higher level of expertise can shorten participants fixations. For the present study longer fixation durations are anticipated for tasks with mixed traffic sign scenario. As members of the treatment group recieved initial introduction materials it can be expected that their mean fixations are lower than those of the control group. Within ethnic groups no differences are estimated.

2.2.4 *Time (Total gaze duration).* Gaze duration in this paper is measured in seconds between the first and last fixation on any region of each task items. So, the summed duration of all fixation on the screen where taken in account. Similar to the mean fixation duration also longer gaze duration indicate more complexity of the stimulus. However, the interpretation exceptions of the previous metric are also valif for gaze duration. For task items where foreign signs are involved and for members of the control group longer total gaze duration is expected. Also, longer time is anticipated on those tas items where the number of traffic signs is highest. The comparison of task type is not intended as pre-tests have showed that solving matching tasks takes generally longer than search or true/false tasks.

3. METHODOLOGY

For a proper implementation and realisation of the eye tracking tests, recommendations implied by [Pernice and Nielsen 2009] were considered.

3.1 Apparatus



(a) Austrian lab: Tobii X50 *(b) USA lab: Tobii 1750*

Fig. 1. Technical infrastructure for the experiment

The data collection for this study was carried out in two laboratories. The data collection for the Austrian participants took place in the video lab at Vienna University of Technology in Austria (Institute of Software Technology and Interactive Systems, Interactive media systems group). The recording of the eye movements of American and Chinese participants was carried out in the eye tracking lab at Clemson University in South Carolina, USA (School of Computing, Visual Computing Group).

To ensure reliable as well as similar technical framework at both locations Tobii eye trackers (the Tobii 50 series) were utilized (see Figure 1). In Austria the Tobii X50 stand alone tracker was used in combination with an IBM 20-inch TFT monitor, whereas in the USA the Tobii 1750 was used, where the eye tracker is integrated in a 17-inch TFT display . The resolution of the screen was set in both labs to 1280x1024 pixels. Both eye trackers are binocular and are equipped with two infrared cameras to capture the coronal reflection of the eye at a sampling rate of 50 Hz. In both labs no chin rest was applied, as freedom of head movement had to be enabled. However, the limitation of participants' head movements was in both technical settings within a range of 30x16x20cm at about 60-70cm from the tracker. Participants were seated generally between 60-72cm from the screen. The accuracy of the eye tracking systems was about 0.5 degrees for the Austrian and 0.5-0.7 degrees for the device in the USA. The spatial resolution was 0.35 for the Tobii X50 and 0.25 for the Tobii 1750 by a potential drift of <1 degrees and a latency of 35 ms for both equipments. Eye movements were recorded from both eyes and for calculation of the fixation the average of both eyes was applied. The calculation of fixations and saccades based on filters defined by [Salvucci and Goldberg 2000]. The velocity threshold was set to 50, whereas a duration threshold of 100 was applied.

The technical setup in Austria was additionally equipped with a live gaze viewer enabling the investigator to simultaneously follow the eye movements on a second monitor during the experiment. Furthermore the data collection in Austria also included recording of audio and video capturing of the participants' upper body. Due to strict regulations of the IRB there was no audio and video recording for American and Chinese participants in the USA.

For gaze analysis and statistical evaluation Tobii Studio (version 2.2.8), R (build 2.11.1) as well as Microsoft Excel (version 2007) were utilized.

3.2 Stimulus

(a) TUWEL course in German *(b) TUWEL course in Chinese* *(c) TUWEL course in English*

Fig. 2. Technical infrastructure for the experiment

The stimuli of this study were international traffic signs from Austria, China and the United States of America embedded into an online course of TUWEL (see tuwel.tuwien.ac.at). TUWEL is the e-learning environment of Vienna University of Technology and it is based on Moodle (version 2.2.1), an open-source community-based e-learning and content management system (see moodle.org). TUWEL has been established as the central e-learning platform during summer term of 2006. Currently, TUWEL serves 30.000+ users within more than 450 courses, dealing with an average

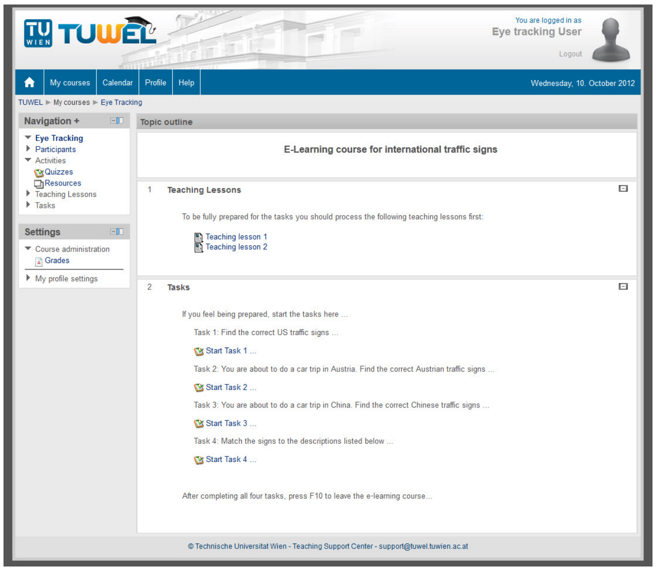
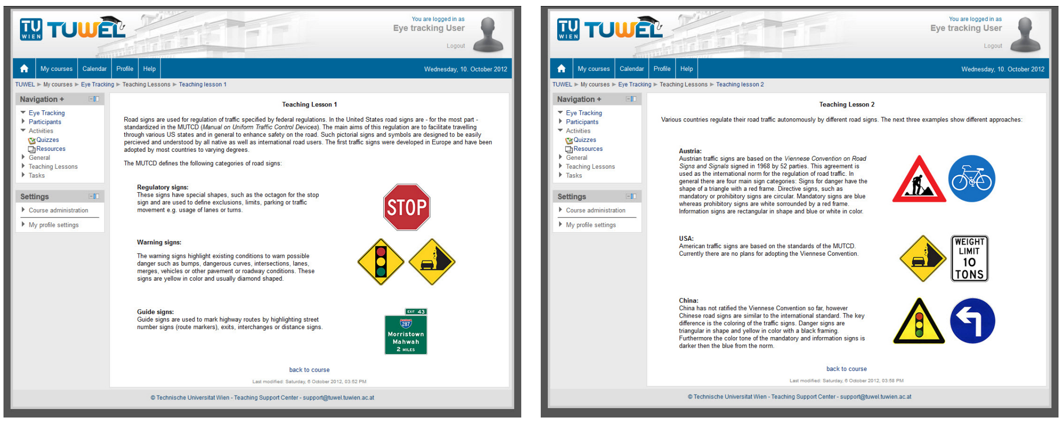


Fig. 3. TUWEL: course page



(a) TUWEL teaching lesson 1

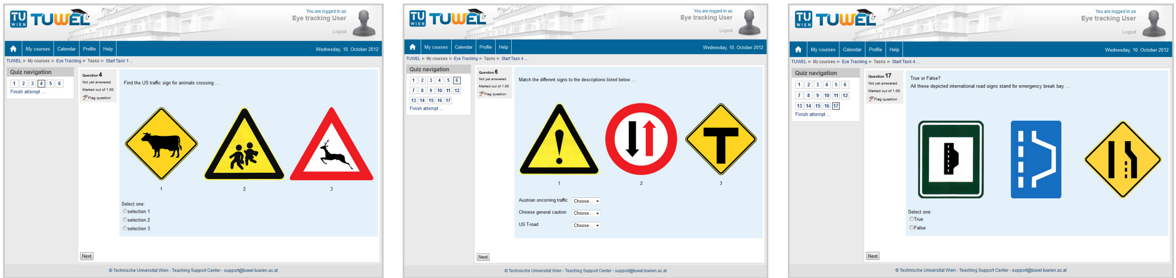
(b) TUWEL teaching lesson 2

Fig. 4. TUWEL: teaching lessons

of more than 6.000 logins per day. The design and graphical user interface of TUWEL is a special development of the Teaching Support Center an institute of the Vienna University of Technology.

In order to minimize language as a factor within the investigation the language of the course as well as the language of the e-learning environment (e.g. navigation menus or buttons) were translated to the official languages of the three countries (see Figure 2). Austrian participants saw the stimulus in German (see Figure 2(a)), volunteers from China saw the e-learning course in Chinese (see Figure 2(b)) and TUWEL was displayed for American users in English (see Figure 2(c)). To be able to compare the eye tracking recordings among the different ethnic groups, the layout of the course as well as of its elements were identically displayed to all participants.

The structure of the e-learning topic consisted of 3 topics (see Figure 3).



(a) *Search task* (b) *Matching task* (c) *True/false task*

Fig. 5. Three different task types used in TUWEL

In topic 0 the title of the e-learning course was displayed informing the learner about the main context of the course.

In topic 1 the two teaching lessons were shown (see Figure 4). Both teaching lessons related to the topic of the investigation serving as online instructions for the oncoming tasks. Teaching lesson 1 (see Figure 4(a)) included general textual information about the domestic signs summarizing hereby the various traffic sign categories and its shapes and coloring. As an visual example for each category a corresponding traffic sign was presented. The second teaching lesson (see Figure 4(b)) contained three paragraphs, one for each of the three countries. The paragraphs included a short description of the main differences of the traffic sign, pointing out differences in shape or coloring. Each paragraph was illustrated jointly by two visuals serving as representative traffic signs for the the specific country.

In topic 2 the main four tasks of the experiment were shown (see Figure 5). Task 1 to 3 were search tasks (see Figure 5(a)), where participants had to find the corresponding traffic signs indicated within the task instructions. In each of these tasks the participants had to find the traffic sign of one specific country. The cover story for the tasks were explained as that the participants are about to travel to the specific country. Each task consisted of 6 items, in total 18 questions had to be answered. To overcome learning effect the order of the tasks was counterbalanced via Latin square. Moreover also the order of the questions (within each task) was displayed randomly to the user. The last task (Task 4) consisted of one search task, eight matching tasks (see Figure 5(b)) and eight true/false-tasks (see Figure 5(c)). Within matching tasks participants had to match different domestic and foreign traffic signs to short textual descriptions. Within the true/false-tasks participants had to make a true/false decision about a textual expression describing a single sign.

The topic of traffic sign was choosen, because of its visual and iconic dimensions that serve as a solid base for investigating visual patterns. The study also represent a testing scenario, where user's knowledge about international traffic signs are surveyed. For test items traffic signs of three countries were chosen: Austria, China and United States of America. Austrian traffic signs are based on the Viennese Convention on Road Signs and Signals signed in 1968 by 52 parties [1]. This agreement is used as the international norm for the regulation of road traffic. China has not ratified the Viennese Convention so far, however Chinese road signs are similar to the international standard. The key difference is the coloring of the traffic signs, whereas shapes and style are adopted from the norm. In contrast, American traffic signs are based on the standards of the MUTCD (Manual on Uniform Traffic Control Devices) and currently there are no plans for adopting the Viennese Convention. Simply put, Austrian signs are closest to the international norm, Chinese road signs are also similar, however having some differences in coloring and at last, American signs have the least similarity to the norm, differing at color, shape and style. For compairison see Figure 6.

The following variations and limitations had been considered when creating the stimuli:

- **Task type**: As mentioned before there are three different task types within the online learning course: (1) search tasks (2) matching tasks (3) true/false tasks

[1] See in Vienna Convention on Road Signs and Signals

Fig. 6. Overview of international traffic signs

- **Sign origin**: Within the study each task type has either questions displaying only domestic signs to the participant or a mixture of domestic and foreign signs from all three countries.
- **Level of reality**: Traffic signs are not shown in real world scenarios, but rather they are depicted as standalone elements reminding the participant taking a driving school's test.
- **Number of images**: For search tasks and matching tasks the number of visuals were both 3 and 5. For true/false tasks the number of images was either 1 or 3.
- **Style of pictograms**: Traffic signs selection for each test item was also randomized on the style of pictogram displayed within the signs. There were two categories defined within this study: (1) 'iconic' pictograms, such as the jumping deer in the wild animal crossing sign, or (2) 'abstract' pictograms (e.g. curved line depicts a dangerous oncoming bend)
- **Complexity of traffic sign:**: Traffic signs selection for each test item was also randomized on the level of complexity. There were two categories defined within this study: (1) simple signs (These are road signs without any pictograms or any simple iconic pictograms.) and (2) complex signs (These are road signs which consist of two or more pictograms.)
- **Sign category**: Traffic signs are split up in 6+1 categories based on the Vienna convention on Road Signs and Signals [2], the international acknowledged system for traffic signs: warning signs (A), priority signs (B), prohibitory signs (C), mandatory signs (D), information signs (F), direction signs (G) and additional panels (H). The last category is excluded due to its solely textual form, as it may evolve language difficulties during the investigation. To ensure that all categories are equally represented in the study traffic signs were used equally distributed in all four tasks.
- **Text within traffic sign**: Selection of traffic signs were based on the criteria that only some of the three countries' road signs might have some short textual elements on them. The majority of the signs however, were not allowed to have any text displayed.

[2]See in Vienna Convention on Road Signs and Signals

- **Reviewing disabled**: Reviewing online tests' results after taking a test is standard procedure. However, in order to avoid learning effects, reviewing test results was disabled for task 1, 2 and 3 and merely enabled for the task 4.

3.3 Procedure

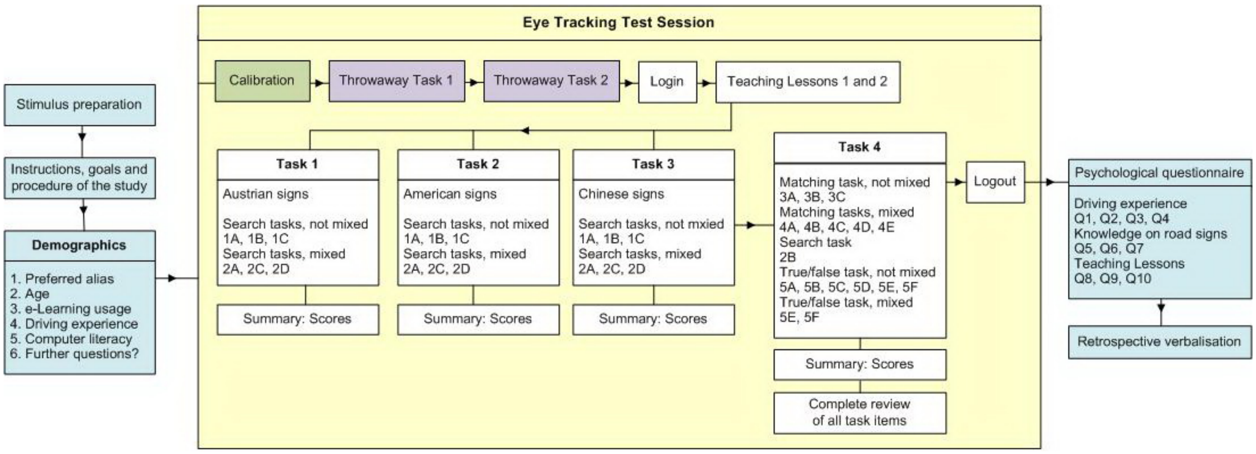
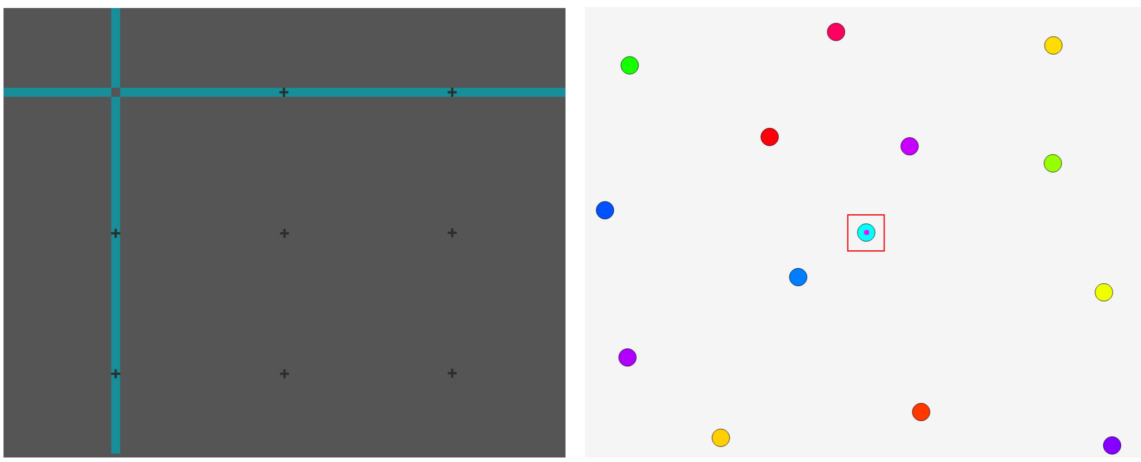


Fig. 7. Workflow diagram of the study



(a) *Calibration grid* (b) *Verification grid*

Fig. 8. Calibration grid and verification grid

The experiment was designed as a single-user test scenario.

having a 3x2x2x2 mixed factorial design, with 2 between-subjects factors (ethnicity at 3 levels: Austria, Chinese, American participants; e-learning instuctions at 2 levels: given or absent) and 2 within-subjects factors (sign origin at 2 levels: foreign or domestic traffic signs; and number of images, e.g., at 2 levels, 3 or 5 images, but only in the search and matching tasks). The study design was blocked by task, where there were 6 tasks in all, 3 types of tasks (search

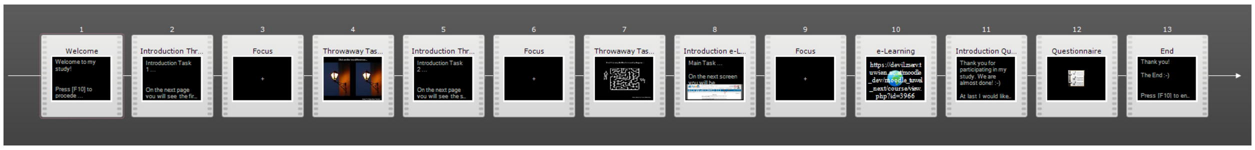


Fig. 9. Stimuli arrangement in Tobii Studio

tasks, matching tasks and true/false tasks) each being homogeneous or mixed in terms of sign origin.

The investigator itself was a non-participating observer sitting behind the volunteers during the test. In Figure 7 the complete work flow of the test is graphically shown.

The first step was the preparation of the stimulus and the verification of the functionality of the technical equipment. After welcoming the participants the goals and the procedure of the test were presented. Three key instructions were told. First, that the experiment is self-paced and participants had no restrictions on the total time of the investigation. Second, volunteers were asked to read the task instructions carefully and solve the tasks as time efficient as possible. To minimize the effects of the intervening personal factor on work accuracy, participants were asked not to hurry, but also not to excessively overthink every task item. Third, participants were told that ’thinking aloud’ is not a criteria for the study, however they can verbally comment any disturbing or affective events during the experiment. Afterwards these initial instructions a pre-questioning on the participants’ demographics was carried out.

As the next step participants were seated as conveniently as possible on a fixed, non-moving chair in front of the monitor with the eye-tracker cameras positioned on their head. Participants were calibrated manually with Tobii’s build in 9-point grid (see Figure 8(a)). Calibration results were validated with Tobii’ integrated verification grid (see Figure 8(b)).

After calibration the eye tracking experiment was started as depicted in Figure 7. The arrangement of visual stimuli in Tobii for instructions, images and other elements can be seen in Figure 9). First participants had to process two throwaway tasks aimed to support users to get used to the test scenario. Hereby focusing elements in form of a small cross were applied to ensure that participants have similar starting points at the beginning of their visual exploration. After completing the throwaway tasks they were asked to login to the e-learning course anonymously with the ’Eye Tracking User’ account displayed within the instruction. After login participants of the control group could read both teaching lessons before completing the four tasks. The order of completion was predefined by the experiment’s design by enumerating the tasks. However, users could freely decide to process the task in an individual order. Among the 36 participants only 2 did not follow the pre-defined order, starting both completing that task, where user were asked to find domestic signs. For task description see previous section 3.2. After task 1,2 and 3 simple summaries with task score are provided, whereas after task 4 not only the scores were presented, but also a complete review of all task items. Immediately after completion of all tasks participants had to log out of the e-learning environment and were asked to complete a questionnaire about their driving experience in Austria, China, United States of America and in foreign countries in general (see questions Q1 to Q4 in Table I). Possible answers were as follows: ”‘Daily”’, ”‘Weekly”’, ”‘Monthly”’, ”‘Quarterly”’, ”‘1-2 times a year”’, ”‘From time to time in my life”’ and ”‘Never”’. Also their level of knowledge about domestic and foreign traffic signs was evaluated, with the following possible answers ”‘Yes, all”’, ”‘More than the half”’, ”‘Approx. the half”’, ”‘Less than the half”’ and ”‘None”’ (see questions Q5 to Q7 in Table I). At last participants’ personal opinions on the teaching lessons’ complexity, difficulty and usefulness was evaluated. Possible answers ramged from ”‘1”’ (corresponding to strong approval) to ”‘7”’ (corresponds to strong disapproval). The questionnaire was designed as a 7-scale Likert scales.

At last, immediately after the eye tracking a session a retrospective interview was carried out, where users’ experiences were jointly reflected with the investigator for about 30 minutes. The interview involved discussions about traffic signs in general, issue concerning tasks, usability problems and task solving strategies. Test sessions’ total duration was about one hour including including pre-questioning, eye tracking and the retrospective analysis.

Q1	Driving experience	How often do you drive a car in Austria?
Q2	Driving experience	How often do you drive a car in the United States?
Q3	Driving experience	How often do you drive a car in China?
Q4	Driving experience	How often do you drive a car in a foreign country in general?
Q5	Level of knowledge	Have you known the Austrian road signs before?
Q6	Level of knowledge	Have you known the US road signs before?
Q7	Level of knowledge	Have you known the Chinese road signs before?
Q8	Learning instructions	If applicable: How simple or complex were the tasks to solve?
Q9	Learning instructions	If applicable: Were the teaching lessons easy to understand?
Q10	Learning instructions	If applicable: Were the teaching lessons useful for solving the tasks?

Table I. Questionnaire's items

3.4 Participants

For this eye tracking study 36 participants were recruited mainly by mail and word-of-mouth recommendations. Some American participants were recruited in a University class, receiving course credit for participation. All the other volunteers did not receive neither money nor any other reward. Participants included 12 Austrian, 12 Chinese and 12 American students (or also University alumni) aged between 22 and 34. The average age of participants was ($M = 26.3$ years). Austrian volunteers average age was higher ($M_{AT} = 28.2$ years) than those of Chinese and American participants, who had almost the same means ($M_{CN} = 25.2$ years) and ($M_{US} = 25.4$ years). To ensure gender balance for each country 6 males and 6 females were recruited. The average years of driving experience was ($M = 7.4$ years). Austrian participants were the most experienced ($M_{AT} = 10.2$ years), followed by the American users ($M_{US} = 8.3$ years) and the Chinese participants, who had the lowest driving experience ($M_{CN} = 3.7$ years).

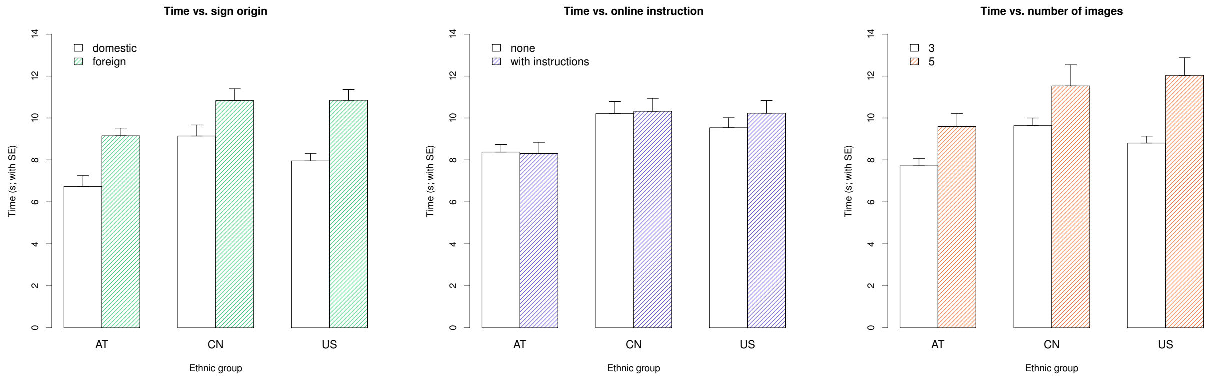
To be accepted to the experiment, one had to meet all the following exclusionary demands. First, participants had to have a driving license (category A and/or B) or a driving permit for at least 2 years in their home countries. Secondly, they had to have some e-learning experience and were not allowed to have high level practical experience with TUWEL. Moreover, all participants had to be at least regular computer users, having basic computer literacy. As known from prior eye tracking tests (see [Nielsen and Pernice 2010]), University freshmen - or novice software users in general - have significantly different visual behavior, so they would generate varying gaze data. At last, participants were not allowed to have any major visual constraints or any other remarkable constraints concerning their physical condition during the eye tracking session. Due to this restriction two volunteers had to be declined, as they could not be calibrated to the eye tracker, because of special eye conditions. For five participants (three Austrian and two Chinese) the eye glasses' reflections impaired the tracking process. As a solution the overhead lights in laboratories where turned off to minimize additional reflections on the surface of the glass. During one experiment a fatal error in Tobii occurred and interrupted the recording. As the participant has already seen the stimulus the experiment had to be aborted and excluded from the results.

As Chinese participants had to be recruited in the USA, they had to meet additional requirements. Only those Chinese volunteers were recruited for the study, whose residency does not exceed one year in the US and having none (or minimal) driving experience on American roads.

Participants were separated in two groups. The control group received no teaching lessons, whereas members of the treatment group were displayed two online instruction within the e-learning course. The selection of the participants was done randomly, however to keep ethnicity and gender in balance 3 males and 3 females of each country were assigned to the control group.

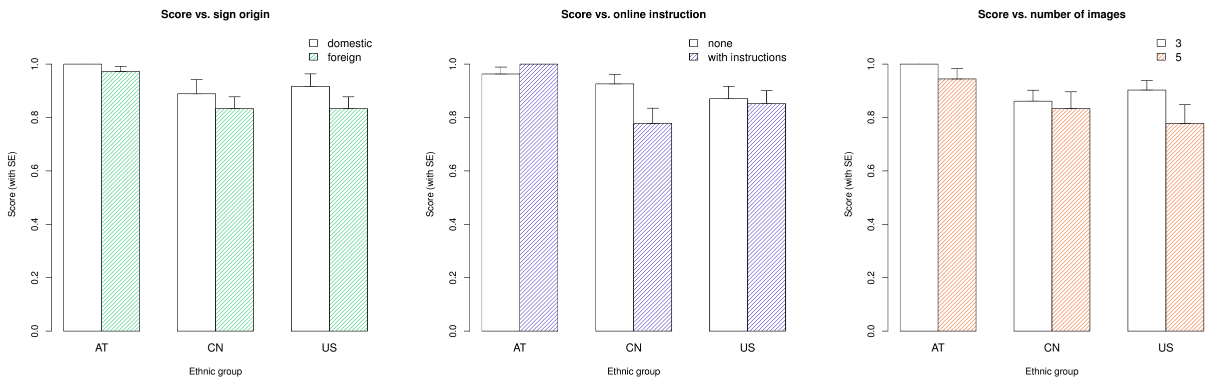
4. RESULTS

Statistical analyses were conducted using *R* [Baron and Li 2007].



(a) effect of sign origin *(b) effect of e-learning* *(c) effect of number of images*

Fig. 10. Search task, homogeneous design: time to task completion.



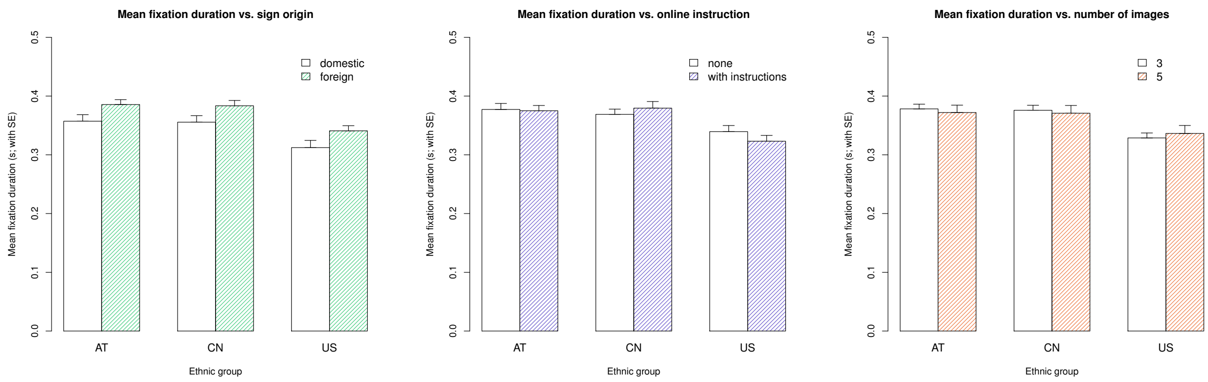
(a) effect of sign origin *(b) effect of e-learning* *(c) effect of number of images*

Fig. 11. Search task, homogeneous design: task success rate.

4.1 Search Task

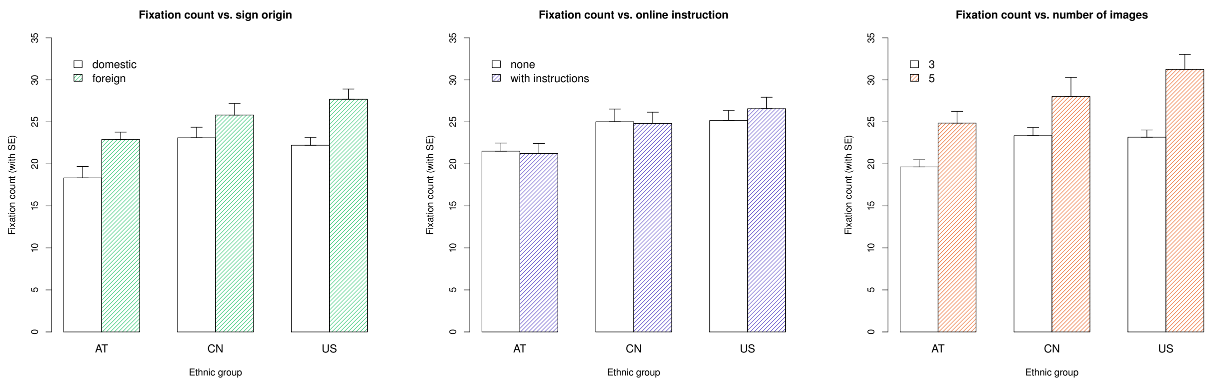
4.1.1 *Homogeneous traffic signs* . When traffic signs (three or five signs in the search field) originated from the same country, a mixed design four-way ANOVA with two between-subjects factors (ethnicity and presence of online instruction) and two repeated measures (traffic sign origin and number of images in the search field) revealed participants' ethnicity as a significant factor on search time completion ($F(2,33) = 5.55, p < 0.01$). Chinese participants were slowest ($M = 10.3$ s) compared to American ($M = 9.9$ s) and Austrian ($M = 8.3$ s) participants, who were fastest. Pairwise comparisons using t-tests with pooled SD and Bonferroni correction shows a significant difference in time to completion between Austrians and Chinese ($p < 0.01$) and a marginally significant difference between Austrians and Americans ($p < 0.05$). No difference was observed between the Chinese and Americans.

The 2×2 mixed factor ANOVA also revealed significant effects of sign origin ($F(1,35) = 30.71, p < 0.01$), number of images ($F(1,35) = 49.60, p < 0.01$), but not of online instruction ($F(1,34) = 0.25, p = 0.62$, n.s.). Two-way interaction between sign origin and number of images was significant ($F(1,35) = 10.27, p < 0.01$) as was the interac-



(a) effect of sign origin *(b) effect of e-learning* *(c) effect of number of images*

Fig. 12. Search task, homogeneous design: mean fixation duration.



(a) effect of sign origin *(b) effect of e-learning* *(c) effect of number of images*

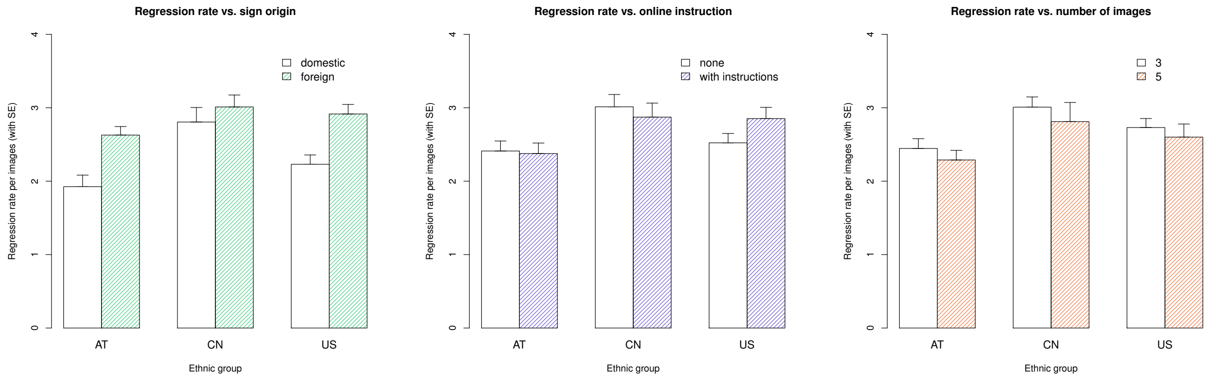
Fig. 13. Search task, homogeneous design: Fixation count.

tion between online instruction and number of images ($F(1,35) = 6.52, p < 0.05$) There was a three-way interaction between ethnicity, sign origin, and number of images ($F(2,34) = 8.57, p < 0.01$).

Effects of sign origin, e-learning, and number of images is shown graphically in Figure 10.

Focusing on task success rate participants' ethnicity also shows significance ($F(2,33) = 4.65, p < 0.05$). The results show that Austrian participants scored highest ($M = 98,2$ %). American ($M = 86,1$ %) and Chinese ($M = 85,2$ %) participants achieved roughly similar results, however having lower success rates than the Austrians. Pairwise comparison with t-tests show significant differences between Austrians and Chinese ($p < 0.01$) and also among Austrians and Americans ($p < 0.01$). As for time to completion there is also no difference between Chinese and American participants for task success ratio.

The 2×2 mixed factor ANOVA revealed a marginal significant effects of sign origin ($F(1,35) = 5.37, p < 0.05$) and number of images ($F(1,35) = 6.50, p < 0.05$). Again, there was no significant effect of online instruction ($F(1,34) = 1.24, p = 0.27$, n.s.). Interaction effect between sign origin and number of images was significant ($F(1,35) = 8.96, p < 0.01$). The three-way interaction between the ethnic groups, sign origin and number of im-



(a) effect of sign origin *(b) effect of e-learning* *(c) effect of number of images*

Fig. 14. Search task, homogeneous design: regression rate on images.

ages was marginally significant ($F(2,34) = 4.29, p < 0.05$). The interaction effect between online instruction and number of images was not significant, however showing tendencies of being so ($F(1,35) = 3.50, p = 0.07$, n.s.). Figure 11 visualizes these effects of sign origin, e-learning, and number of images.

The eye tracking parameter *mean fixation duration* has no significance for ethnicity ($F(2,33)=2.29, p=0.11$, n.s.), as all ethnic groups show roughly the same results ($M_{AT} = 0.3761$ s), ($M_{CN} = 0.3742$ s), ($M_{US} = 0.3313$ s). The pairwise comparison with t-tests indicate no significant differences among the ethnic groups. However, the 2×2 mixed factor ANOVA reveals a strong significance for the factor of sign origin ($F(1,35)=18.71, p<0.01$), but not for the factors number of images ($F(1,35)=0.05, p=0.83$, n.s.), online instruction ($F(1,34)=0.01, p=0.89$, n.s.), or any of the two-way interactions. The three-way interaction between the ethnic groups, sign origin and number of images was not significant either, but shows tendencies to have an effect ($F(2,34)=3.15, p=0.06$, n.s.). Figure 12 visualizes the effects of sign origin, e-learning, and number of image on mean fixation duration.

The next parameter the overall fixation count has marginal significance for ethnicity ($F(2,33)=3.98, p<0.05$) as Austrian participants seam to need less fixation for viewing traffic signs ($M=21,38$ fixations) than the Chinese ($M=24.92$ fixations) or the American participants ($M = 25.87$ fixations). Similar to score or time pairwise comparison with t-tests show significant differences among Austians and Chinese ($p<0.05$) as well as Austrians and Americans ($p<0.01$), however there are no differences between the Chinese and American ethnic groups.

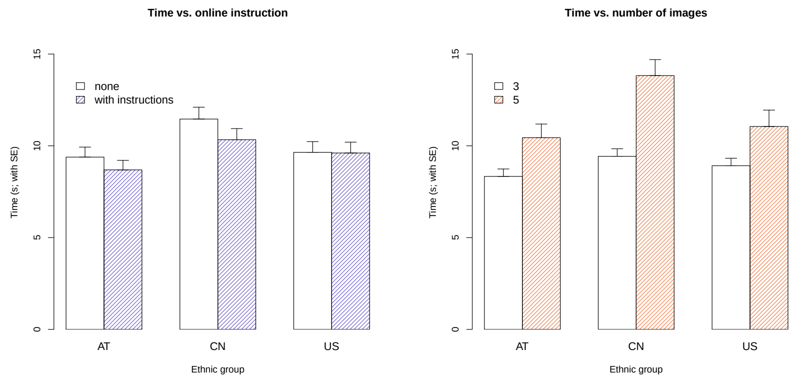
Investigating the factors there is again a high significance for number of images ($F(1,35) = 55.75, p < 0.01$) and sign origin ($F(1,35)=20.80, p<0.01$). For the last factor online instruction there are no significant differences shown ($F(1,34)=0.05, p=0.82$, n.s.).

Focusing on the two-way interactions between the factors the following effects can be reported. There are marginal significant results for the interaction between online instruction and number of images ($F(1,35)=5.58, p<0.05$) as well as sign origin and number of images ($F(1,35) = 6.87, p < 0.05$). The three-way interaction between the ethnic groups, sign origin and number of images was highly significant again ($F(2,34) = 9.35, p < 0.01$). Effects of sign origin, e-learning, and number of images is shown graphically in Figure 13.

The ANOVA analysis reveals that the ethnicity of the participants have a marginal significant effect on the regression rate of images ($F(2,33)=3.92, p<0.05$). Austrian participants have less average regressions on images ($M=2.394$ regressions) than American ($M=2.687$ regressions) or Chinese participants ($M=2.943$ regressions).

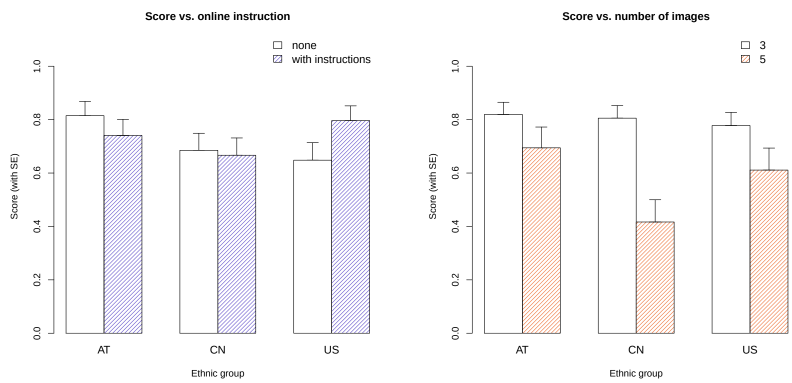
The 2×2 mixed factor ANOVA also revealed high significant effects for sign origin ($F(1,35)=21.61, p<0.01$). In contrast to this. there are no significant differences shown for number of images ($F(1,35)=3.67, p=0.065$, n.s.) or online instruction ($F(1,34)=0.10, p=0.75$, n.s.).

Analysis of two-way interaction among factors show no overall significant differences, however the three-way interaction effect yields a highly significant result for ethnicity, sign origin and number of images ($F(2,34)=7.89, p<0.01$). Effects of these factors on the regression rate can be seen in Figure 14.



(a) effect of e-learning *(b) effect of number of images*

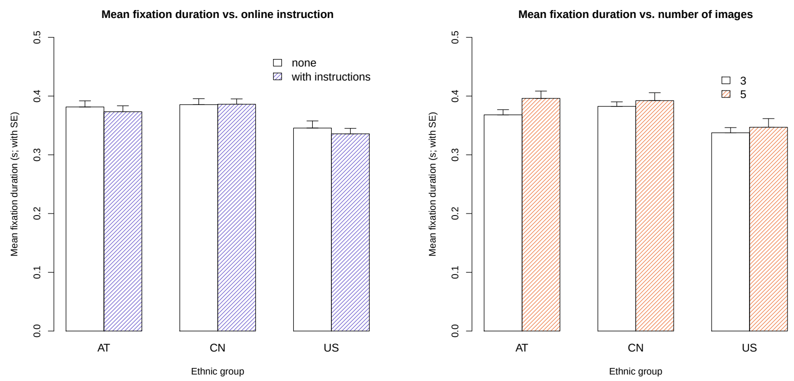
Fig. 15. Search task, mixed design: time to task completion.



(a) effect of e-learning *(b) effect of number of images*

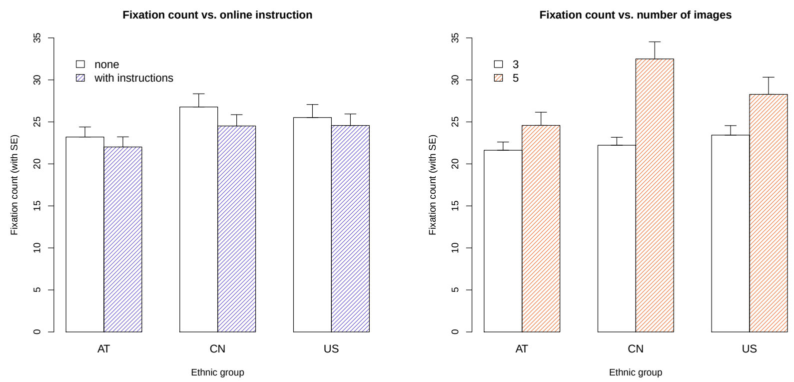
Fig. 16. Search task, mixed design: task success rate.

4.1.2 *Mixed traffic signs.* When traffic signs (three or five signs in the search field) originated from three different countries (Austria, China and United States of America), a mixed design four-way ANOVA with two between-subjects factors (ethnicity and presence of online instruction) and one repeated measure (number of images in the search field) revealed participants' ethnicity as a marginally significant factor on search time completion ($F(2,33)=3.93, p<0.05$). Austrian participants were the fastest ($M_{AT}=9.037$ s), followed by American ($M_{US}=9.628$ s) and the Chinese participants ($M_{CN}=10.89$ s). Pairwise comparison with t-tests reveal statistical significance only between Austrian and Chinese participants ($p<0.01$), in contrast there were no differences among other ethnic group combinations.



(a) effect of e-learning *(b) effect of number of images*

Fig. 17. Search task, mixed design: mean fixation duration.



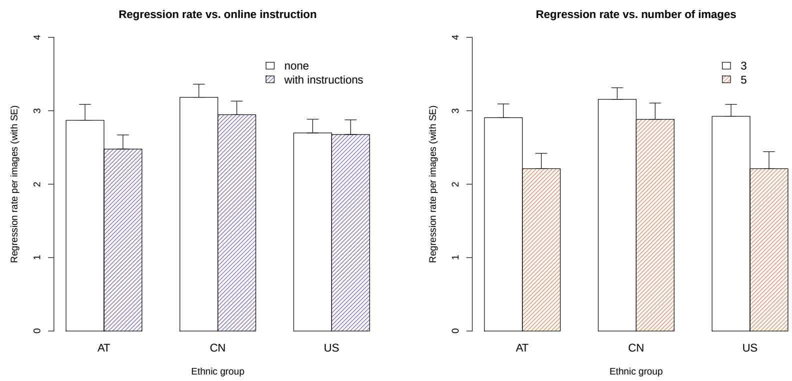
(a) effect of e-learning *(b) effect of number of images*

Fig. 18. Search task, mixed design: fixation count.

The ANOVA analysis further showed that for mixed traffic signs the number of images is highly significant ($F(1,35) = 37.30, p < 0.01$). No significant differences were found for the online instruction ($F(1,34) = 1.25, p = 0.27$, n.s.) , all two-way and the three-way interaction effects. In Figure 15 the effects of e-learning and number of images are shown graphically.

The task success rate has no significant effect for ethnicity ($F(2,33) = 0.85, p = 0.43$, n.s.), as the results of all the three involved ethnic groups were similar ($M_{AT} = 77.8$ %), ($M_{US} = 67.6$ %) ($M_{CN} = 72.2$ %). High statistical significance were found for number of images ($F(1,35) = 18.67, p < 0.01$). Similar to time no statistically significant results could be revealed for the online instruction ($F(1,34) = 0.09, p = 0.77$, n.s.), nor for any of the two-way or the three-way interaction combinations of selected factors. Effects of e-learning and number of images are shown graphically in Figure 16.

There are no significant effects of the ethnicity on the mean fixation duration factor ($F(2,33) = 2.15, p = 0.13$, n.s.), as participants of all three ethnic groups have similar average values ($M_{AT} = 0.3774$ s), ($M_{US} = 0.3406$ s) ($M_{CN} = 0.3857$ s). The pairwise t-tests indicate however that there is statistical difference between Austrian and Amer-



(a) effect of e-learning *(b) effect of number of images*

Fig. 19. Search task, mixed design: regression rate on images.

ican as well as between American and Chinese participants (both $p < 0.01$). The ANOVA analysis shows marginal significance effects of number of images ($F(1, 35) = 4.37, p < 0.05$), but there were no differences detected for the online interaction factor ($F(1, 34) = 0.09, p = 0.76$, n.s.) nor for any of the two-way or the three-way interaction combinations of selected factors. In Figure 17 the effects of e-learning and the number of images is depicted.

Participants' ethnicity is not a statistically significant factor on the eye tracking parameter *fixation count* ($F(2, 33) = 1.82, p = 0.18$, n.s.). There is also no significant effect on the online instruction factor ($F(1, 34) = 1.14, p = 0.29$, n.s.) Also the pairwise t-tests show no differences among the ethnic groups. However, the ANOVA analysis reveal high difference for *number of images* ($F(1, 35) = 24.59, p < 0.01$) and indicate statistical tendencies for the two-way interaction effect between ethnic groups and number of images ($F(2, 33) = 3.25, p = 0.053$, n.s.), where Chinese and American participants have larger differences between tasks having 3 or 5 images. This effect can be clearly seen in image (b) of Figure 18. The three-way interaction between ethnicity, online instructions and number of images was marginally significant ($F(2, 34) = 3.4, p < 0.05$). Effects of e-learning and number of images are shown graphically in Figure 18.

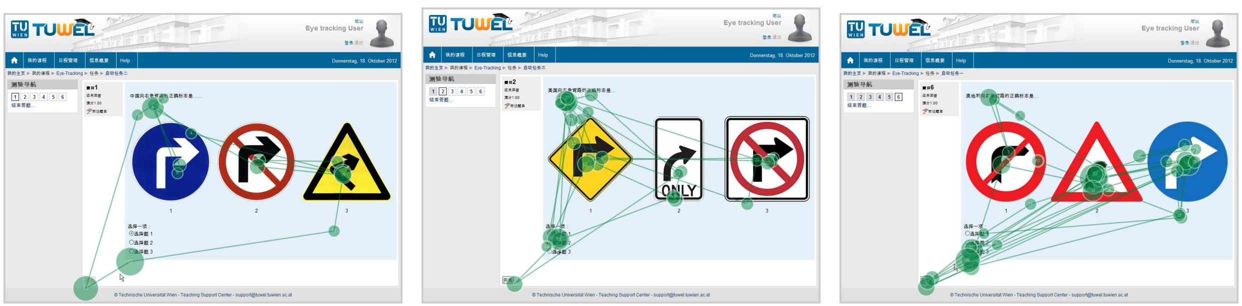
The eye tracking metric *regression rate* is not statistically significant for ethnicity ($F(1, 34) = 1.28, p = 0.29$, n.s.), neither show the pairwise t-tests any differences among the ethnic groups, as they have similar regression rates ($M_{AT} = 2.674$ regressions), ($M_{US} = 2.687$ regressions) ($M_{CN} = 3.065$ regressions). Furthermore neither online instruction ($F(1, 34) = 0.91, p = 0.34$, n.s.), nor the two-way combinations have any statistical significance. According to the ANOVA results there are again high differences for number of images ($F(1, 35) = 14.59, p < 0.01$) and marginal significant effects for the three-way interaction between ethnic groups, online instruction and number of images ($F(2, 34) = 4.81, p < 0.05$). In Figure 19 the effects of e-learning and the number of images are depicted.

4.1.3 *Discussion.* For both scenarios—mixed and homogeneous traffic signs—the number of images has strongest influence. A higher number of images within tasks increases not only time to task completion, but also has a clear effect on the eye tracking metric fixation count. Furthermore, the analysis reveals that for a higher number of images (for both mixed and homogeneous scenarios) participants' means fixation durations are similar. Considering that no two-way interaction was observed, the effect between ethnicity and number of images indicates that participants cope similarly with tasks with three or five images, regardless of their origin. Simply put, more images require longer visual inspection.

More interesting, perhaps, is that this similar effect can be seen for the regression rate. Participants' rate of re-visiting images within search tasks seems to be independent of the number of images. Again, as there is no interaction effect between ethnicity and number of images, it can be stated that members of all cultures perceived both search tasks' visual stimuli in a similar way.

Perhaps surprisingly, there appears to be no effect of online instruction on any of the metrics investigated within this study. Members of the treatment group (those seeing teaching materials prior to the test) have no divergent eye movement patterns in either search task. Also there are no overall differences in time to task completion or overall success rate. Only the two-way interaction effect between online instruction and number of images indicates some significance, which however is more likely caused by a differing number of images. Statistical means suggest in various cases that teaching materials could even have a negative effect on the metrics, e.g., increasing time to task completion and fixation count, and impairing task success rate. This overall effect will be further discussed in Section 5.

Interpretation of the ANOVA results also indicates a strong overall influence of sign origin for all metrics. Foreign signs appeared to yield significantly longer times to task completion, lower scores, more fixations, higher regression rates, and longer mean fixation durations. Overall, interpretation of this effect would be that participants exhibit higher cognitive load during perception of foreign signs. Figure 20 shows an example, where a Chinese participant needs significantly more fixations and regressions for the task that depicts foreign signs only (see Figure 20(a)). The two-way interaction effect suggests there are no differences in terms of ethnicity, meaning that all three ethnic groups had difficulty fixating foreign signs. There is an effect between sign origin and number of images for all metrics expect mean fixation duration and regression rate. However, this is again more likely caused by the number of images.



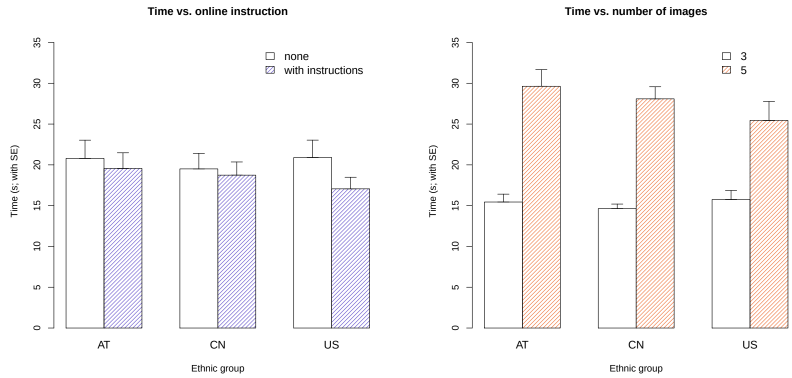
(a) domestic signs (Chinese) *(b) foreign signs (American)* *(c) foreign signs (Austrian)*

Fig. 20. Gaze plot comparison of visual search on domestic and foreign traffic signs

According to ANOVA, ethnicity seems to have a divergent effect on the metrics of both search tasks. While results for the homogeneous scenario (foreign and domestic signs displayed separately) show significant differences for nearly all metrics, there are hardly any divergent results for the mixed scenario. There are significant differences only in time to completion for both scenarios, as in both cases Austrian participants were significantly faster, followed by the American and Chinese volunteers, the latter being the slowest. Score and eye movement patterns were not divergent in the mixed scenario, however they do differ significantly in the homogeneous scenario, where Austrian participants showed higher scores, the lowest fixation counts and the lowest regression rates. In contrast, American users however, showed lowest mean fixation durations. Furthermore, as there are no two-way interactions between ethnicity and any other factor, it seems all ethnic groups perform similarly among different numbers of images or online instructions.

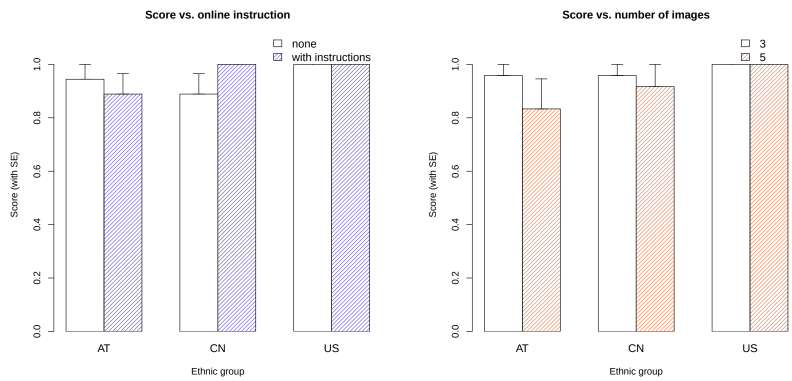
4.2 Matching Task

4.2.1 *Homogeneous traffic signs.* When traffic signs (three or five signs in a matching text-to-image scenario) originated from the same country, a mixed design four-way ANOVA with two between-subjects factors (ethnicity and presence of online instruction) and one repeated measure (number of images in the matching field) revealed participants' ethnicity not to be a significant factor on time to completion ($F(2,33) = 0.27, p = 0.76$, n.s.). Also the



(a) effect of e-learning *(b) effect of number of images*

Fig. 21. Matching task, homogeneous design: time to task completion.



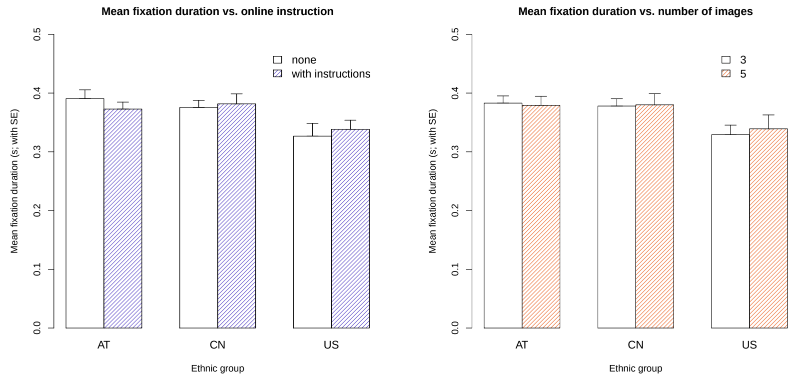
(a) effect of e-learning *(b) effect of number of images*

Fig. 22. Matching task, homogeneous design: task success rate.

pairwise t-tests show no significance among the ethnic groups, as they have roughly similar results for time ($M_{AT} = 20.18$ s), ($M_{US} = 18.98$ s) and ($M_{CN} = 19.12$ s). Moreover neither the online instruction ($F(1, 34) = 1.83, p = 0.18$, n.s.), nor any of the two-way or three-way interactions among the factors show significant differences. For time to task completion only number of images show highly significant effects ($F(1, 35) = 181.27, p < 0.01$). Effects of e-learning and number of images are shown graphically in Figure 21.

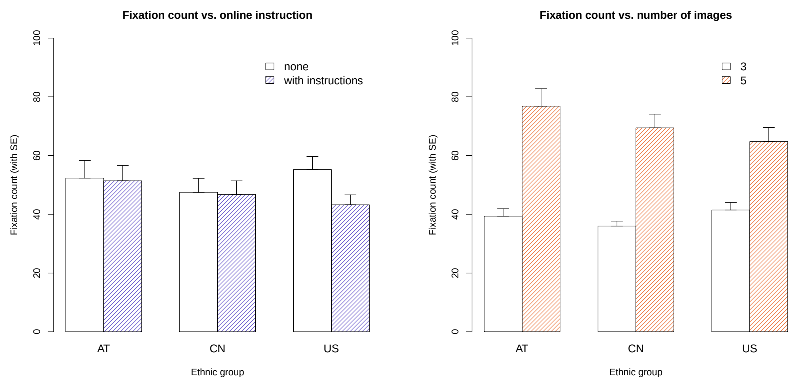
Focusing on scores none of the factors, ethnic groups or two-way or three-way interactions among factors show significant effects. Also the pairwise tests show no significant differences among the ethnic groups, as all of them performed similarly well during these domestic matching tasks, see ($M_{AT} = 91.7$ %), ($M_{CN} = 94.4$ %) and Americans even ($M_{US} = 100$ %). The ANOVA results are for ethnicity ($F(2, 33) = 1.66, p = 0.21$, n.s.), for online instructions ($F(1, 34) = 0.23, p = 0.63$, n.s.), and for number of images ($F(2, 33) = 1.67, p = 0.21$, n.s.). Summaries of e-learning and number of images effects are shown graphically in Figure 22.

Similar to task success rate, none of the factors, ethnic groups or two-way interactions among the factors show significant effects, only the three-way interaction between ethnic group, online instruction and number of images have



(a) effect of e-learning *(b) effect of number of images*

Fig. 23. Matching task, homogeneous design: mean fixation duration.

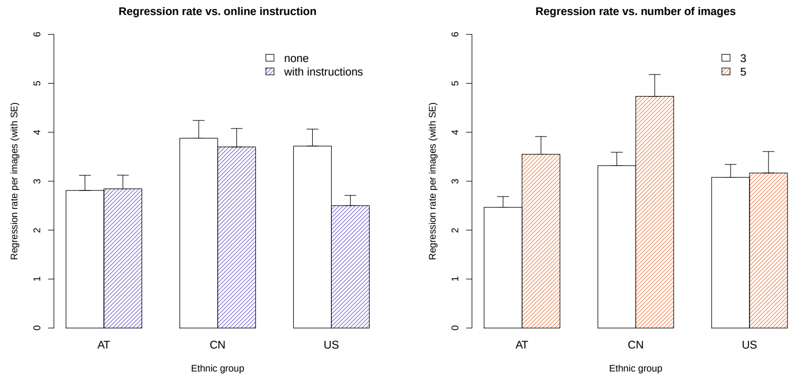


(a) effect of e-learning *(b) effect of number of images*

Fig. 24. Matching task, homogeneous design: fixation count.

marginal significance ($F(2, 34) = 3.67, p < 0.05$). Furthermore the pairwise t-tests show some differences among the ethnic groups. Austrian ($M_{AT} = 0.3817$ s) and Chinese participants' mean fixation durations ($M_{CN} = 0.3786$ s) were significantly higher than those of the American participants ($M_{US} = 0.3325$ s). ANOVA results confirm these effects with ($p < 0.01$) for both cases. Between the Chinese and Austrian participants' mean fixation durations there are no statistical differences detected. The results for the other non-significant factors are as follow. Ethnic groups results are ($F(2, 33) = 2.24, p = 0.23$, n.s.), online instructions results are ($F(1, 34) = 1.67, p = 1.00$, n.s.) and for number of images the results are ($F(1, 35) = 0.21, p = 0.65$, n.s.). In Figure 23 the effects of e-learning and the number of images are graphically shown.

There is no significant effect of ethnicity on fixation count ($F(2, 33) = 0.58p = 0.56$, n.s.). Also the pairwise t-tests show no differences among the ethnic groups, as the groups fixation counts are circa on similar levels. Chinese participants fixation count is the lowest ($M_{CN} = 47.14$ fixations), followed by the Americans ($M_{US} = 49.22$ fixations) and the Austrians, having the highest fixation count among all countries ($M_{AT} = 51.86$ fixations). The ANOVA analysis shows that there is no significant difference for online instruction ($F(1, 34) = 1.64, p = 0.21$, n.s.). However,



(a) effect of e-learning *(b) effect of number of images*

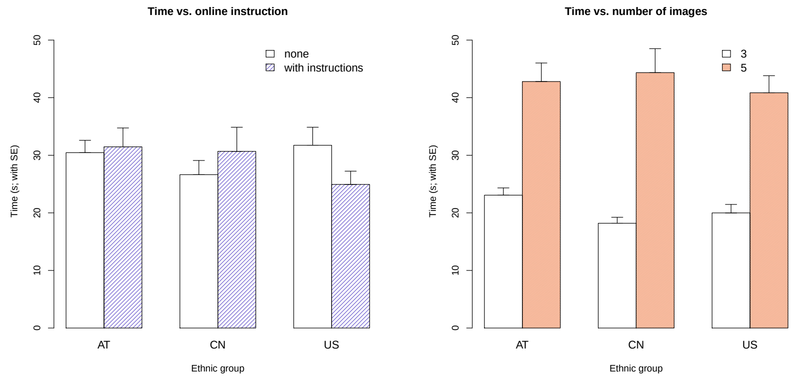
Fig. 25. Matching task, homogeneous design: regression rate on images.

there is highly significant effect for number of images on the fixation count metric ($F(2,34)=0.67, p<0.01$). Also, there is a statistical tendency of the two-way interaction between ethnic groups and number of images ($F(2,34)=3.25, p=0.052$, n.s.). At last, there is no statistical difference for the three-way interaction of all three factors. Effects of e-learning, and number of images is shown graphically in Figure 24.

The ANOVA analysis reveals no statistical significance for ethnicity on participants' regression rates ($F(2,33)=2.76, p=0.079$, n.s.). Austrian participants have the lowest regression results ($M_{AT}=2.828$ regressions), followed by the Americans ($M_{US}=0.3.1.08$ regressions) and the Chinese participants needing the most average re-visits to images ($M_{CN}=3.789$ regressions). With the pairwise t-tests it can be seen that there are marginal significant differences among the Austrian and the Chinese participants ($p<0.05$). There are no statistically significant differences among other ethnic groups. The analysis for online instruction reveals no significant influence ($F(1,34)=1.74, p=0.19$, n.s.), neither does the three-way interaction between all factors. However, there is a highly significant effect on number of images ($F(1,35)=20.77, p<0.01$) as well as a marginal effect on the two-way interaction between ethnic group and number of images ($F(2,34)=4.45, p<0.05$). This effect can be clearly seen in image (b) of Figure 25. Effects of e-learning and number of images are graphically depicted in Figure 25.

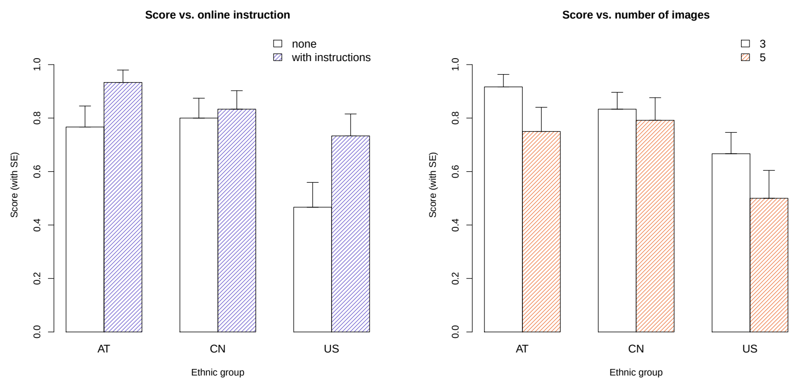
4.2.2 *Mixed traffic signs.* When traffic signs (three or five signs in a matching text-to-image scenario) originated from three different countries, a mixed design four-way ANOVA with two between-subjects factors (ethnicity and presence of online instruction) and one repeated measure (number of images in the matching field) revealed participants' ethnicity not to be a significant factor on time to task completion ($F(2,33)=0.38, p=0.68$, n.s.). Austrian participants were the slowest ($M_{AT}=30.96$ s), followed by Chinese ($M_{CN}=28.66$ s) and the American participants, who were the fastest ($M_{US}=28.34$ s). Pairwise t-tests reveal, however no significance among these ethnic groups. Also online instruction has any significant effect on time ($F(1,34)=0.04, p=0.83$, n.s.) as have all the two-way and three-way interaction among the factors. High statistical significance has number of images on time to task completion ($F(1,35)=130.14, p<0.01$), as participants needed more than twice the time to solve tasks with 5 images than those featuring 3 images, ($M_3=20.42$ s) compared to ($M_5=42.67$ s). Effects of e-learning, and number of images are summarized in Figure 26.

In contrast to time ethnicity has high statistical influence on task success rate of participants of each group ($F(2,33)=7.01, p<0.01$). Participants of Austria achieved the highest score ($M_{AT}=85$ %), followed by the Chinese ($M_{CN}=81.7$ %) and the American participants ($M_{US}=60$ %). Pairwise t-tests indicate high differences between Austrian and the American users ($p<0.01$) and marginal significance between Chinese and American participants ($p<0.05$).



(a) effect of e-learning *(b) effect of number of images*

Fig. 26. Matching task, mixed design: time to task completion.

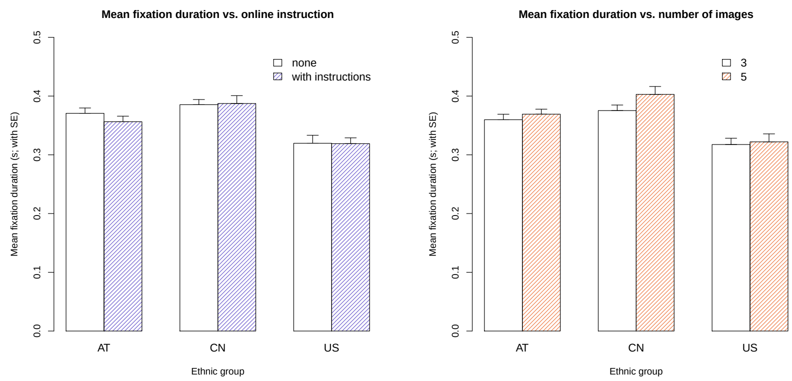


(a) effect of e-learning *(b) effect of number of images*

Fig. 27. Matching task, mixed design: task success rate.

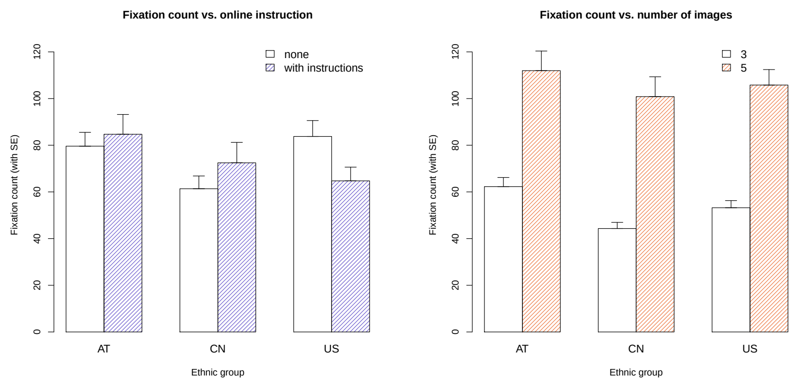
Furthermore ANOVA analysis reveals marginal significant differences for the online instruction ($F(1,34)=6.90, p<0.05$) as well as for number of images ($F(1,35)=6.59, p<0.05$). The results show no significant difference for any of the two-way or three-way interactions between factors. Effects of e-learning and number of images are graphically depicted in Figure 27.

The ANOVA analysis reveals that there is a marginally significant effect of ethnicity on participants' the mean fixation durations ($F(2,33)=4.71, p<0.05$). Chinese participants have the longest mean fixation durations ($M_{CN}=0.3863$ s), followed by the Austrians ($M_{AT}=0.3635$ s) and the US participants having again the lowest fixation duration values ($M_{US}=0.3193$ s). Pairwise t-tests show high significant differences between the Austrian and American and also between the Chinese and the American users (both $p<0.01$). No significance was revealed between the Austrian and Chinese participants. Focusing on number of images the analysis indicates high significance of this factor on the mean fixation durations ($F(1,35)=8.71, p<0.01$). In contrast to this online instruction ($F(1,34)=0.05, p=0.81$, n.s.), as both the two-way and three-way interactions between the factors show any significant differences. In Figure 28 the effects of e-learning and the number of images are graphically shown.



(a) effect of e-learning *(b) effect of number of images*

Fig. 28. Matching task, mixed design: mean fixation duration.

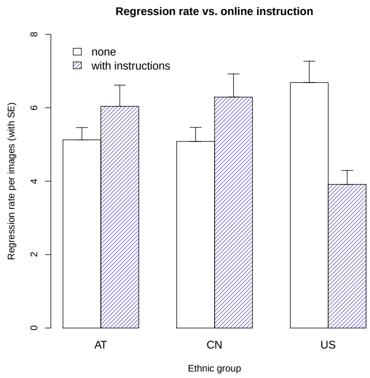


(a) effect of e-learning *(b) effect of number of images*

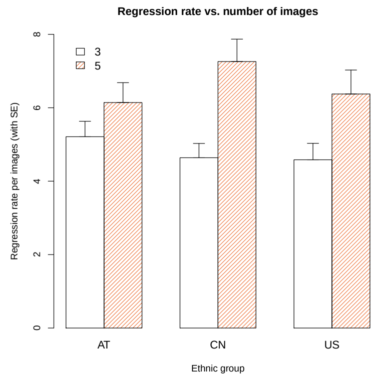
Fig. 29. Matching task, mixed design: fixation count.

For the eye tracking metric *fixation count* there was no significant difference revealed in terms of ethnicity ($F(2,33)=2.42, p=0.11$, n.s.). The Austrian users had the highest fixation count ($M_{AT}=82.15$ fixations) compared to the American ($M_{US}=74.25$ fixations) and to the Chinese participants ($M_{CN}=66.92$ fixations), who needed the least amount of fixations during task completion. Pairwise t-tests reveal no significant effects among the ethnic groups. Online instruction does also have no significant effect on fixation count ($F(1,34)=0.03, p=0.86$, n.s.), as do the two-way and three-way interactions between the factors. For fixation count only number of images shows high statistically significant differences ($F(1,35) = 161.59, p < 0.01$). Participants need for tasks displaying five images roughly twice as much fixations as for tasks featuring three images, ($M_3=53.27$ fixations) compared to ($M_5=106.2$ fixations). Effects of e-learning and number of images are shown graphically in Figure 29.

The ANOVA analysis reveals no significant differences for ethnicity on participants' regression rates ($F(2,33) = 0.15, p = 0.86$, n.s.), as there were no differences shown for online instruction ($F(1,34) = 0.13, p = 0.71$, n.s.). Focusing on the regression rates, it can be clearly seen that all ethnic groups have similar results on this eye tracking parameter: ($M_{AT} = 5.583$ regressions), ($M_{CN} = 5.687$ regressions) ($M_{US} = 5.3$ regressions). Pairwise t-test show



(a) *effect of e-learning*



(b) *effect of number of images*

Fig. 30. Matching task, mixed design: regression rate.

analog to the ANOVA results no differences among the ethnic groups. For three-way interaction effects between factors there were no significance shown, however there is a marginal difference for the two-way interaction between ethnic groups and online instruction ($F(2,33)=4.63, p<0.05$). Mean tables reveal that American participants having online instructions prior to solving the tasks need significantly less regressions than users from Austria or China. This effect can be clearly seen in image (a) of Figure 30. Focusing on number of images again, there is high significance for this factor on regression rate ($F(1,35) = 36.19, p < 0.01$). Tasks with three images need less re-visits ($M_3 = 4.811$ regressions) than those displaying five images ($M_5 = 6.592$ regressions). Effects of e-learning and number of images are shown graphically in Figure 30.

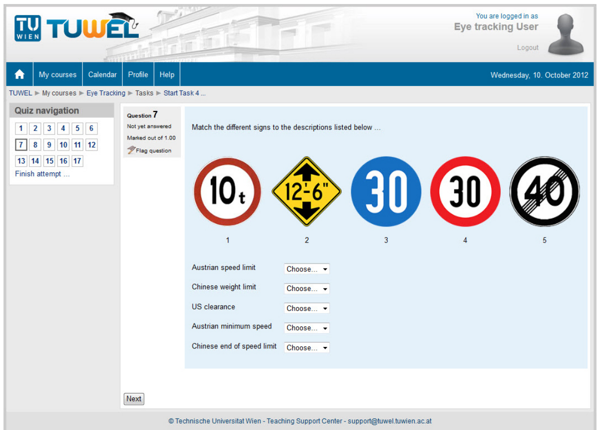
4.2.3 *Discussion.* Similar to search tasks, for both—homogeneous and mixed task designs—online instruction has again no overall effect on any parameters, except on task success rate for the mixed scenario. In this special case American participants (those of the treatment group) seem to significantly benefit from perceiving teaching lessons. The two-way ANOVA results on the interaction effect between online instruction and regression rate indicate to have an evidence for this. American participants' re-visits on foreign visual stimuli significantly decrease if they get prior online instructions. No similar effect could be seen for Austrian or Chinese participants.

For matching tasks number of images has again the strongest influence on all metrics. More images within a matching task significantly increase—for both homogeneous and mixed design—time to task completion, fixation count and regression rate. Also the task success rate drops in case of higher number of images, however this effect in not statistically significant.

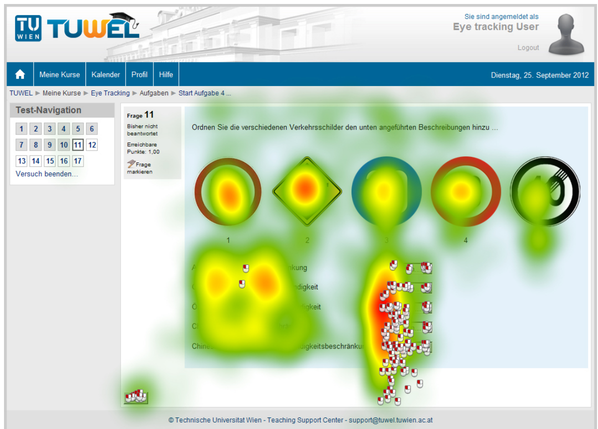
Mixed scenarios simultaneously depicting domestic and foreign traffic signs provoke higher mean fixation durations in case of higher number of images. This effect can be seen as participants exhibit higher cognitive load. Also participants qualitative feedback goes along with this effect, as many volunteers reported difficulty with solving matching tasks depicting mixed traffic signs. As there is no two-way effect between mean fixation duration and ethnic groups or online instruction, it can be stated that all ethnic groups have similar visual patterns in this particular case.

Ethnicity in general has no significant effect on matching tasks. Participants eye tracking metrics are similar regardless of their ethnic origin. However there is again, one significant exception. Interestingly similar to search tasks, American participants have again significantly lower mean fixation durations than Austrian or Chinese volunteers, as pairwise t-tests reveal. This effect can be seen for both—homogeneous and mixed task design—, however it is only significant for the mixed scenario. If considering task success rates this outcome yields some interesting results. American participants achieve higher scores with lower mean fixation durations than the other two ethnic groups.

However, this statement is only correct for the homogeneous design. For mixed scenario task success ratio is exactly the other way round, as Americans achieve the lowest score with the lowest mean fixation duration. Simply put, it means that Americans visually inspect matching tasks at higher speed (dynamics). For visual elements they know, they achieve good results efficiently, however for unfamiliar items they are in favor over missing on details (CHECK THIS SENTENCE).



(a) Matching task, mixed traffic signs



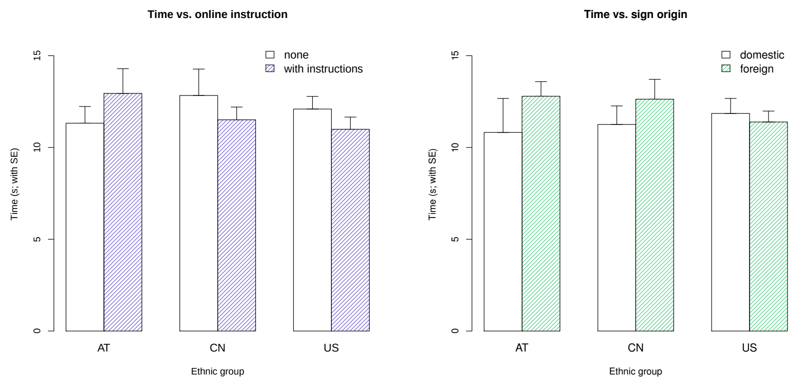
(b) Austrian participant's heatmap

(c) Chinese participant's heatmap

(d) American participant's heatmap

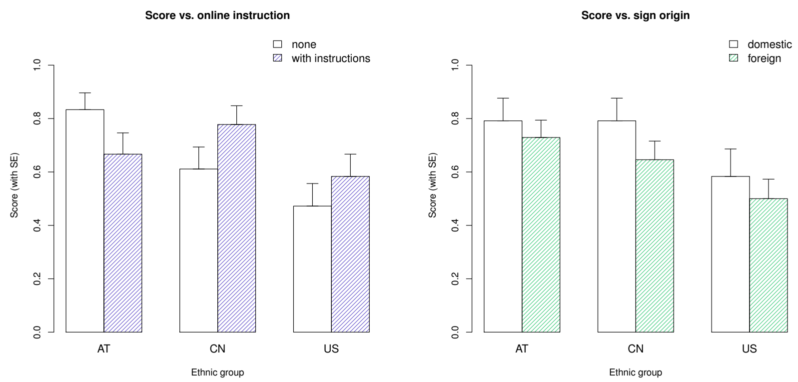
Fig. 31. Fixation intensities are higher on foreign traffic signs

Similar to search tasks, foreign traffic signs elicit significantly higher fixation intensities within matching tasks. As an example in Figure 31 heatmaps are displayed for all ethnic groups. Each ethnic group has highest fixation counts on foreign traffic signs. Chinese and American participants visual observe Austria's circular signs with highest intensity (see Figure 31(c) and Figure 31(d), whereas Austrian participants have hot spots on Chinese and American road signs (see Figure 31(b).



(a) effect of e-learning *(b) effect of sign origin*

Fig. 32. True/false task, homogeneous design: time to task completion.



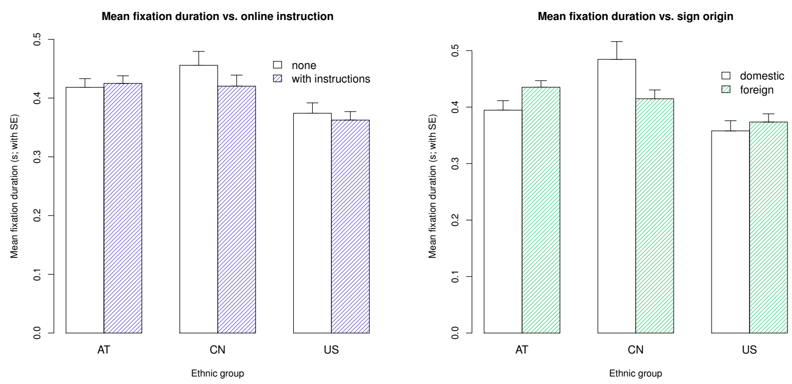
(a) effect of e-learning *(b) effect of sign origin*

Fig. 33. True/false task, homogeneous design: task success rate.

4.3 True/false Task

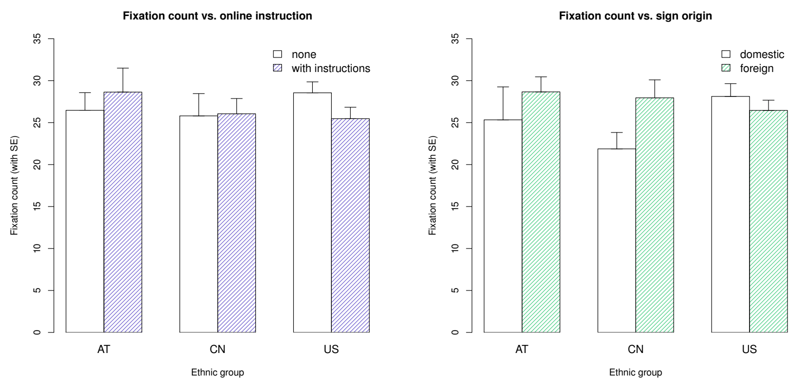
4.3.1 *Homogeneous traffic signs.* When traffic signs in a true/false scenario originated from either domestic or foreign countries, a mixed design four-way ANOVA with two between-subjects factors (ethnicity and presence of online instruction) and one repeated measure (sign origin of the image) revealed no significance for ethnicity on time to task completion ($F(2,33) = 0.16, p = 0.84$, n.s.), as all the ethnic groups have similar time results. American were the fastest ($M_{US} = 11.54$ s), followed by the Austrian participants ($M_{AT} = 12.13$ s) and by the Chinese users ($M_{CN} = 12.17$ s). Pairwise t-test show no differences among the ethnic groups. For both online instruction ($F(1,35) = 0.07, p = 0.79$, n.s.) and for sign origin ($F(1,35) = 1.91, p = 0.18$, n.s.) there were no significant results to be found. Also the two-way and three-way interactions between the factors show no significant differences on time to task completion. Effects of e-learning and sign origin are shown graphically in Figure 32.

In contrast to time, ethnicity shows highly significant differences on task success rate ($F(2,33) = 7.53, p < 0.01$). Austrian users achieved the highest rate ($M_{AT} = 75$ %), followed by the Chinese participants ($M_{CN} = 69.4$ %) and the Americans achieving the lowest results ($M_{US} = 52.78$ %). Pairwise comparisons using t-test reveal marginally



(a) effect of e-learning *(b) effect of sign origin*

Fig. 34. True/false task, homogeneous design: mean fixation duration.

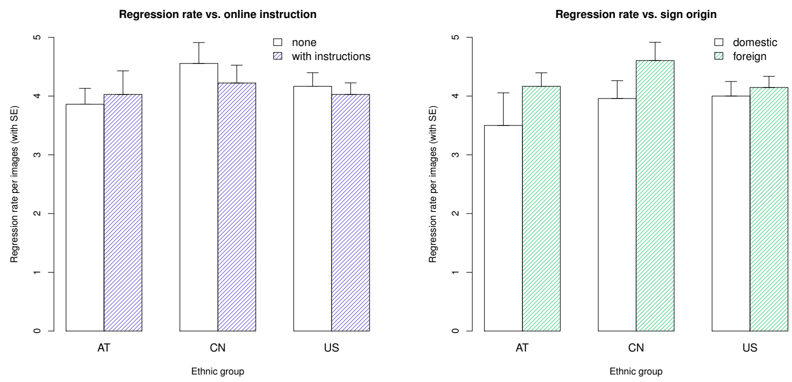


(a) effect of e-learning *(b) effect of sign origin*

Fig. 35. True/false task, homogeneous design: fixation count.

significant differences between Austrian and American participants ($p < 0.05$). No significant effect could be detected between other ethnic group combinations. As ANOVA reveals online instruction shows no significant differences ($F(1,34) = 0.57, p = 0.45$, n.s.), as does sign origin ($F(1,35) = 1.96, p = 0.17$, n.s.). Although there is a marginal significance for the two-way interaction between ethnic group and online instruction ($F(2,33) = 4.49, p < 0.05$). Austrian participants receiving online instructions in the e-learning environment achieve lower scores than those receiving to teaching lessons, ($M_{AT-no} = 83$ %) compared to ($M_{AT-yes} = 67$ %). American and Chinese users perform better when reading teaching instructions prior to the true/false tasks. This effect can be clearly seen in image (a) of Figure 33. Two-way interaction effects involving sign origin and three-way interaction effect between all factors show no significant results. In Figure 33 the effects of e-learning and sign origin are graphically depicted.

The ANOVA analysis reveals that ethnicity has no significant effect on the eye tracking metric *mean fixation duration* ($F(2,33) = 2.34, p = 0.11$, n.s.). Similar to prior results American participants have the lowest mean fixation durations ($M_{US} = 0.3683$ s), followed by the Austrian users ($M_{AT} = 0.4217$ s) and the Chinese participants ($M_{CN} = 0.4381$ s). Pairwise comparisons using t-test indicate high significant differences between Austrian and



(a) effect of e-learning *(b) effect of sign origin*

Fig. 36. True/false task, homogeneous design: regression rate.

American as well as between Chinese and American participants (both $p < 0.01$). No significant effects were shown between the Austrian and Chinese groups. If focusing on online instruction, the ANOVA analysis do not reveal any significant differences here ($F(1, 34) = 0.24, p = 0.62$, n.s.). Also the sign origin does not have any significant effect on participants' mean fixation durations ($F(1, 35) = 0.17, p = 0.68$, n.s.). However, the two-way interaction effect between sign origin and ethnic group reveal a highly significant difference ($F(2, 34) = 9.49, p < 0.01$). The mean values tables reveal that both Austrian and American users have lower mean fixation durations on domestic signs than on foreign traffic signs. Chinese participants however have higher mean fixation values on Chinese signs, ($M_{CN-domestic} = 0.48$ s) compared to ($M_{CN-foreign} = 0.41$ s). Two-way interaction effects involving online instruction and three-way interaction effect between all factors show no significant results. Effects of e-learning and sign origin are shown graphically in Figure 34.

Ethnicity has no significant effect on fixation count ($F(2, 33) = 0.17, p = 0.84$, n.s.), as all ethnic groups have roughly similar values on this eye tracking metric. Chinese users have the lowest fixation count ($M_{CN} = 25.93$ fixations), followed by the American participants ($M_{US} = 27.01$ fixations). Austrian participants had the highest fixation count among all ethnic groups ($M_{AT} = 27.56$ fixations). Pairwise t-test also indicate no significant differences among the ethnic groups. Online instruction has no effect on fixation count ($F(1, 34) = 0.01, p = 0.92$, n.s.). In contrast to sign origin shows statistical tendencies ($F(1, 35) = 3.39, p = 0.07$, n.s.). As the grand mean values reveal fixation counts on domestic signs seem to be lower than on foreign traffic signs, ($M_{domestic} = 25.11$ fixations) compared to ($M_{foreign} = 27.69$ fixations). Two-way and three-way interaction effects among factors show no statistically significant differences on fixation count. Effects of e-learning and sign origin are depicted graphically in Figure 35.

Ethnic groups shows no significant effects on the eye tracking metric *regression rate* ($F(2, 33) = 0.78, p = 0.47$, n.s.), as all the ethnic groups' participants have similar results. Austrian users had the lowest regression rate ($M_{AT} = 3.944$ regressions), tightly followed by the American participants ($M_{US} = 4.097$ regressions) and the Chinese, who had the highest regression rates ($M_{CN} = 4.389$ regressions). Pairwise comparisons using t-test indicate no statistically significant differences among the ethnic groups. However, sign origin reveals a marginally significant effect on users' regression rates ($F(1, 35) = 4.33, p < 0.05$). According to the results domestic traffic signs were perceived with lower regression rates than foreign signs, ($M_{domestic} = 3.819$ regressions) compared to ($M_{foreign} = 4.306$ regressions). All two-way or three-way interaction effects between factors show no significance on the eye tracking metric *regression rate*. In Figure 36 the effects of e-learning and sign origin are graphically depicted.

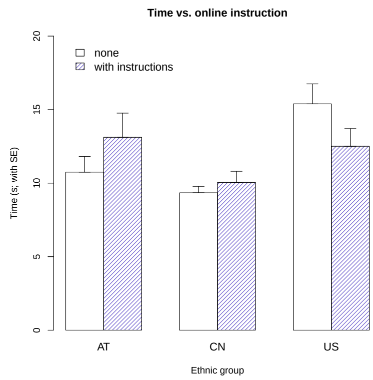


Fig. 37. True/false task, mixed design: time to task completion.

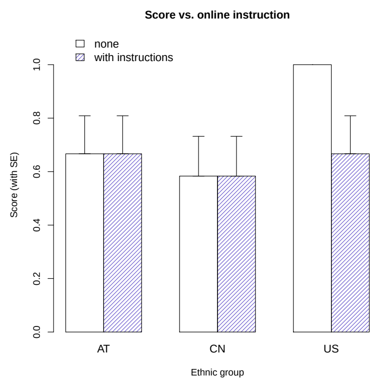


Fig. 38. True/false task, mixed design: task success rate.

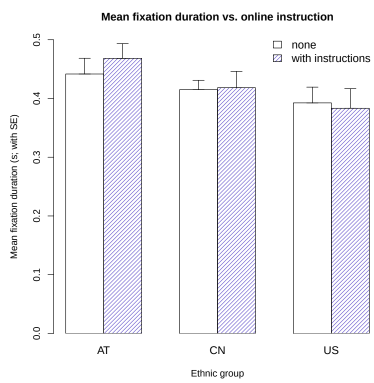


Fig. 39. True/false task, mixed design: mean fixation duration.

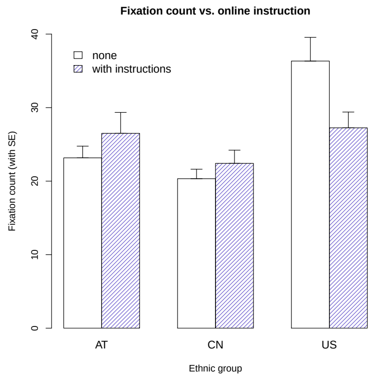


Fig. 40. True/false task, mixed design: fixation count.

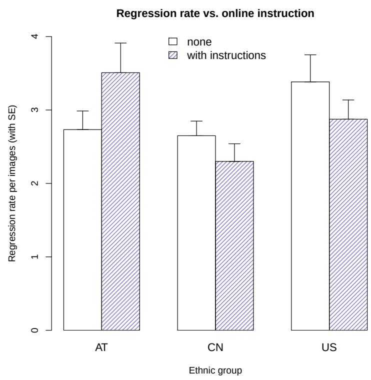


Fig. 41. True/false task, mixed design: regression rate.

4.3.2 *Mixed traffic signs.* When traffic signs in a true/false scenario originated from three different countries (Austria, United States and China), a two-way ANOVA with two between-subjects factors (ethnicity and presence of online instruction) was carried out. The results revealed a highly significant effect of ethnicity on time to task completion ($F(2,69) = 6.89, p < 0.01$). Chinese participants were the fastest ($M_{CN} = 9.696$ s), followed by the Austrian users ($M_{AT} = 11.933$ s) and the American participants ($M_{US} = 13.950$ s). Pairwise combinations using t-test show marginal significance between Chinese and American participants ($p < 0.05$). No significant differences were found between the other ethnic groups. The ANOVA analysis indicates no significant differences for online instruction ($F(1,70) = 0.01, p = 0.94$, n.s.), however there is statistical tendency revealed for the two-way interaction effect between ethnic groups and online instruction ($F(2,69) = 2.74, p = 0.07$, n.s.). The tables of means show that both Austrian and Chinese users need longer, if they received online instruction prior to the tasks. However, American participants seem less time for time completion in case of prior e-learning instructions, ($M_{US-no} = 15.392$ s) compared to ($M_{US-yes} = 12.508$ s). This effect of e-learning is depicted graphically in Figure 37.

Not one factor, nor the two-way interaction between the factors show any statistical significance on task success rate. Ethnicity result is ($F(2,69) = 1.85, p = 0.16$, n.s.). American users had the highest task success rate ($M_{US} = 83.3$ %), followed by the Austrian participants ($M_{AT} = 66,7$ %) and the Chinese users ($M_{CN} = 58.3$ %). Pairwise combinations using t-test show no significant effect between the ethnic groups. Also, the two-way interaction effect between ethnicity and online interaction have no significant effect on score ($F(1,70) = 1.06, p = 0.30$, n.s.). The effect of e-learning is shown graphically in Figure 38.

The ANOVA analysis reveals a marginally significant effect of ethnicity on participants' mean fixation durations ($F(2,69) = 3.23, p < 0.05$). As for other task types, also for true/false tasks (mixed scenario) the American users had the lowest mean fixation duration ($M_{US} = 0.3879$ s), followed by the Chinese ($M_{CN} = 0.4167$ s) and the Austrian participants ($M_{AT} = 0.4550$ s). Furthermore pairwise t-test indicate a marginal significant effect between the Austrian and the American groups ($p < 0.05$). No other significant differences could be detected for the other ethnic group combinations. At last, neither online instruction ($F(1,70) = 0.10, p = 0.75$, n.s.), nor the two-way interaction between the factors reveals any significance ($F(2,69) = 0.23, p = 0.79$, n.s.). In Figure 39 the effect of e-learning is graphically shown.

Ethnicity has a highly significant effect on participants' fixation counts ($F(2,69) = 11.06, p < 0.01$). The Chinese have the lowest value for fixation count ($M_{CN} = 21.38$ fixations), followed by the Austrian users ($M_{AT} = 24.83$ fixations) and the American participants, who had the highest fixation count values among all three ethnic groups ($M_{US} = 31.79$ fixations). Pairwise t-tests reveal a highly significant effect between Chinese and Americans ($p < 0.01$) and a marginal significance between Austrian and American users ($p < 0.05$). There were no significant differences found for online instruction ($F(1,70) = 0.44, p = 0.50$, n.s.), however the two-way interaction effect between ethnic groups and online instruction shows a marginally significant influence on the eye tracking parameter *fixation count* ($F(2,69) = 4.59, p < 0.05$). As depicted in Figure 40 it can be clearly seen, that both Austrian and Chinese users have higher fixation counts, if they have received teaching lessons before taking the test. In contrast to this, those American users have higher fixation counts, who had no online instructions prior the test, ($M_{US-no} = 36.33$ fixations) compared to ($M_{US-yes} = 27.25$ fixations).

Similar to fixation count also for the regression rate there is a marginally significant difference among the ethnic groups ($F(2,69) = 3.19, p < 0.05$). Chinese participants had the lowest regression rates ($M_{CN} = 2.475$ regressions). American and Austrian users had almost the same values for this eye tracking metric, ($M_{AT} = 3.1208$ regressions) compared to ($M_{US} = 3.1292$ regressions). However, pairwise t-tests reveal no statistically significant differences for the ethnic groups. Similar to the previous metrics, online instruction has no significant effect on regression rate ($F(1,70) = 0.01, p = 0.90$, n.s.), neither does the two-way interaction effect between ethnic groups and online instruction ($F(2,69) = 2.77, p = 0.07$, n.s.). The latter merely indicates some statistical tendencies again, that there is a difference between Americans compared to Chinese and Austrian participants in terms of regression rate. Those Americans, who had no teaching lessons at the beginning of the test, seem to have higher regression rates, than those, who received online instruction prior to the test, ($M_{US-no} = 3.383$ regressions) compared to ($M_{US-yes} = 2.875$ regressions). The general effect of e-learning is shown graphically in Figure 41

4.3.3 *Discussion.* Sign origin has again a significant influence on true/false tasks, however the effect size is smaller compared to search tasks. In general, foreign traffic signs increase participants fixation counts and regression rates indicating higher visual exploration. Furthermore sign origin seem to have a two-way interaction effect with ethnicity, as there are some significant differences for score and mean fixation durations among different ethnic groups. As an example again, American volunteers had significantly lowest mean fixation durations for both true/false task scenarios among all ethnic groups. Gazeplots of exemplary participants shown in Figure 42 clearly demonstrate that the sizes of the circles (corresponding to fixation duration) are smaller for American volunteers (see 42(b)) than for Chinese or Austrian participants (see 42(c) and 42(d)). However, it has to be said, there is no significant correlation between this eye tracking metric and task success rate, as analysis reveal.

Probably the most interesting result for ethnicity is the differences between both scenarios. All ethnic groups seem to have similar results for eye tracking metrics for the homogeneous design, however according to ANOVA nearly all metrics have significant differences for the mixed scenario. General patterns indicate that Chinese participants were faster, having fewer fixations and perceiving tasks with lower regressions rates than users of the other ethnic groups.

(a) *True/false task, mixed traffic signs*

(b) *American participant's gazepath*

(c) *Chinese participant's gazepath*

(d) *Austrian participant's gazepath*

Fig. 42. Gazeplots demonstrate lower mean fixation duration for American participant

Also that US perceive visual stimuli with higher fixation intensities and elicit more re-visits in order to re-compare images or its features.

At last, like matching and search tasks online instruction has no overall significant effect on the eye movements. However, there is one interesting and statistically significant effect that has to be discussed. Results yield that members of the American treatment group have higher mean fixation durations, higher fixation counts and higher regression counts than those of the control group. Simply put, it indicates that those Americans, who saw teaching lessons prior to the test, visually perceive true/false tasks (in the mixed scenario) more efficiently. However, if this assumption is correlated with task success rate, it can be clearly seen, that American members of the treatment group have significantly lower scores than those of the control group. Therefore it is more likely that the differences within the eye tracking metrics arise from personal (other psychological) attributes.

4.4 Questionnaire' results

Immadiately after the eye tracking test session a questionnaire was to be completed by all participants. The results are listed in Table II

Austrian participants have highest driving experience on Austrian roads. Their average of 2.6 indicates monthly driving. American volunteers show highest driving experience on US roads with an average of 1.5 revealing weekly driving. Also Chinese participants have highest driving experience on domestic roads, however their average of 4.5 indicate quarterly driving intensity. Moreover questionnaire's results reveal that Chinese participants have high driving experience on American or foreign roads in general. Participants of the control and treatment group show similar results in term of driving experience.

Focusing on level of knowledge each ethnic groups' representatives show highest level of knowledge on their domestic traffic signs. The level of approval is roughly the same among all countries ($M_{AT} = 1.3$), ($M_{US} = 1.2$) and ($M_{US} = 2.0$), indicating that they know all or more than half of the domestic traffic signs. Interestingly, Chinese participants indicate again high level of knowledge on American signs. The reason for this might be, that Chinese participants are living temporarly in the USA, intend to apply for American driving license and herefor already gained some knowledge about the American traffic signs. However, results clearly show a general low level of knowledge on foreign road signs. Also there are no significant differences between the control and treatment groups' results.

The questionnaries' results on teaching materials reveal that participants consider online instructions rather simple than complex, easy to understand and rather useful than unuseable. Low standard deviation indicate coherency among participants' feedbacks. Also, the ethnic groups' results on learning instructions do not show any differences, indicating that volunteers of all ethnic groups had similar opinion on the teaching materials.

Results	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Mean	5,4	3,7	6,1	5,4	3,8	2,3	3,9	2,9	2,2	2,8
Min	1	1	2	1	1	1	1	1	1	1
Max	7	7	7	7	5	5	5	5	4	6
Stdv	2,3	2,8	1,6	2,1	1,8	1,3	1,5	1,2	1,1	1,5
Median	7	2,5	7	6	5	2	5	3	2	3
Mean-AT	2,6	6,8	7,0	5,7	1,3	3,8	4,8	2,8	2,2	2,8
Mean-US	7,0	1,5	6,9	6,8	5,0	1,2	4,8	3,1	2,3	2,8
Mean-CN	6,6	2,8	4,5	3,8	5,0	1,9	2,0	2,8	2,0	2,9
Mean-control	5,3	3,8	6,0	5,6	3,7	2,3	3,6	x	x	x
Mean-treatment	5,4	3,6	6,3	5,2	3,8	2,3	4,1	3,0	2,2	3,1

Table II. Questionnaire's results

4.5 Qualitativ analysis

-Teaching Lessons 1: —American participants seam to ignore bullet point list more than CN/AT —American F-shape reading pattern, not so CN/AT

-Teaching Lessons 2: —heatmaps reveal highest fixation intensity for foreign countries' descriptions —heatmaps reveal participants read first foreign country's description more intense than the description of the second (positioned in lower areas of the screen) —Americans lower fixation intensities

4.6 Quantitative Results

In this subsection the quantitative results of the investigation are summarized. Table III shows all aggregated values for all the conditions and independend variables.

5. GENERAL DISCUSSION

5.1 Influence of teaching materials

As results show there is hardly any overall effect of teaching materials on participants' eye movements. For the present study no significant differences could be revealed neither for all investigated task types nor for any of the eye tracking metrics. There is only one significant result for the parameter score on matching tasks in the special scenario of mixed international traffic signs. However, in depth analysis of the data reveals that there has been a strong effect for American participants of the treatment group having generous benefit from teaching materials. This had extensive effect on the overall outcome of the dependend variable. Some other tendencies indicate that American volunteers seem to have lower regression rates on foreign signs, if they had teaching lessons prior to the test.

The effect that teaching lessons have influence can be caused by various factors. One reason might be that the quality of the teaching materials was low. However, the questionnaire's results reveal that online instructions were perceived rather simple, useful and easy to understand coherently among participants. An other reason might be that participant's level of knowledge on traffic signs prior to the study was high, as the average of driving experience was 7 years. It can be assumed that volunteers used rather heuristics than gained nwe knowlegde to solve the tasks. This assumption goes along with participants' feedbacks during the retrospective interview, where many of them have reported solving tasks by applying the strategy of excluding traffic signs they don't know. Also the deisgn and layout of tasks might have been the factor to imply application of exclusion, as the test items were strongly visual (depicting more likely pictures instead of textual instructions). Eye movements on the teaching lessons revealed that the online instructions have been read intensively by all ethnic groups. Fixations clusters cover all areas, where important knowledge was described. Fixation intensities were even higher on descriptions of foreign traffic signs. Simply put, participants did not skip the instructions to read, they have even focued on crutial part of the text.

5.2 Influence of sign origin

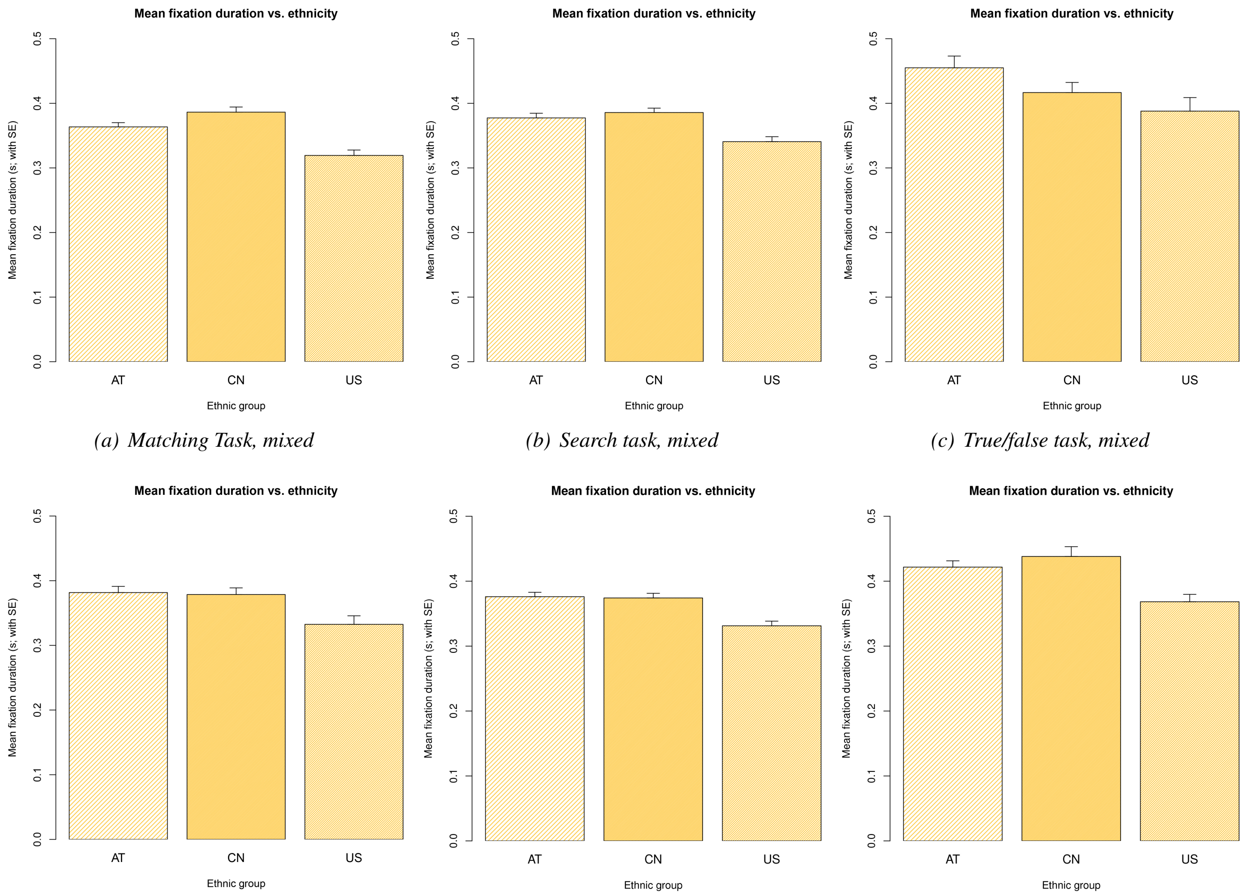
Sign origin had an overall effect for almost all metrics and all task types. The strongest statistical effect was measrued for search tasks, the weakest for true/false tasks. In general foreign signs provoke lower scores, higher fixation duration, higher fixation count, higher mean fixation duration and higher regression rates on traffic signs. Also, mixed scenarios of e-learning tasks, where traffic signs of different countries were displayed at the same time provoked lower scores, higher fixation counts. As reference see Table III. However, both effects were expected, as several publications in eye tracking point out, that unfamiliarity and out-of-context objects generate higher fixation intensities.

5.3 Influence of culture

Focusing on the overall outcome of each dependent variable there are no significant difference to report for gaze duration, fixation count and regression rate. However there is statistical evidence for different mean fixation durations for American participants compared to Chinese and Austrians volunteers. Overall in the presented study American's mean fixation durations were the lowest, for every blocking task types and task types' scenarios, as depicted in Figure 43. This results seams to follow findings of [Chua et al. 2005] indicating that Americans tries to fixate faster on the foreground objects in order to retrieve most visual details. As American participants fixation count is highest among along ethnic groups, results indicate that the willingness to spend adequate time on tasks excludes the psychological factor of impatience. Due to the facts that Austrian participants had highest scores and driving experience also the impact of expertise can be excluded. Possible reason for this effect might be explained by faster extraction of visual information or differences in motivation or concentration levels.

	Gaze duration (time in sec)			**Score** (points)			**Mean fixation** duration (sec)			**Fixation count** (number of fixations)			**Regression rate** (integer)		
Overall	15.3			0.78			0.379			37.5			3.6		
Search tasks (mixed)	9.8			0.72			0.367			24.4			2.8		
Search tasks (homogeneous)	9.5			0.89			0.360			24.1			2.6		
Matching tasks (mixed)	19.4			0.75			0.356			74.4			5.5		
Matching tasks (homogeneous)	29.3			0.95			0.364			49.4			3.2		
True/false tasks (mixed)	11.8			0.69			0.419			26			2.9		
True/false tasks (homogeneous)	11.9			0.65			0.409			26.8			4.1		
Group (treatment / control)	15.1	15.6		0.78	0.77		0.378	0.381		36.8	38.2		3.45	3.6	
Ethnicity (Austria / USA / China)	15.4	15.5	15.1	0.82	0.74	0.76	0.395	0.346	0.396	38.3	38.8	35.3	3.4	3.6	3.6
Scenario (mixed / homogeneous)	13.7	16.9		0.72	0.83		0.381	0.378		41.6	33.4		3.7	3.3	
Sign origin (domestic / foreign)	9.6	11.2		0.82	0.75		0.377	0.388		23.1	26.6		3.1	3.6	
Number of images (1/3/5)	8.8	14.6	35.2	0.86	0.80	0.79	0.361	0.363	0.365	22.2	37.2	88.3	2.8	3.2	5.2

Table III. Quantitative Re- view



(a) Matching Task, mixed *(b) Search task, mixed* *(c) True/false task, mixed*

(d) Matching Task, homogenous *(e) Search task, homogenous* *(f) True/false task, homogenous*

Fig. 43. Mean fixation durations are lower for American participants on all task types and scenarios.

When compairing different task types, results reveal an interesting outcome. For matching tasks and true false tasks significant differences were found for the mixed scenario, suggesting that there are differences in visual perception among ethnic groups, when fixating both domestic and foreign signs within one task item. However, this effect seems to be inversed for search tasks, where homogeneous presentation of traffic signs yield to cause differences in eye movements. Simply put, all ethnic groups have similarly visual percpetion on traffic signs only if they are homogeneoulsy displayed in matching or true/false tasks or are embedded heterogeneously (mixed) in search tasks. However, cultural differences in eye movements could derive from several sources, including differences in socialization or physical state of condition, such as motivation or concentration., aspects which were however not insluded within the present study.

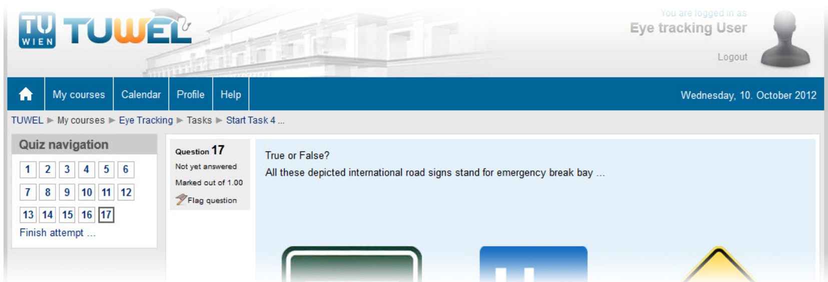
5.4 Influence of number of images

number of images ANOVA revealed strongest effect for number of images. Not a surprising effect, however its dimensions are.

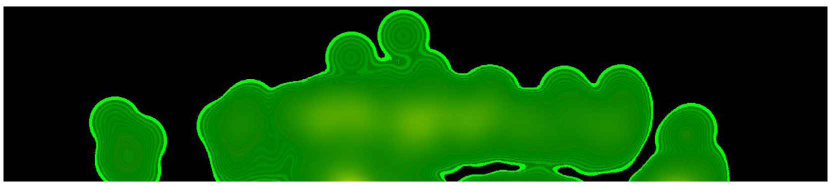
Idee - number of images strongest effect - no na eh klar - aber spannend ist der ausma, statt mal 1,4 ist der effekt oft mal 2 (nachrechnen)!!!

5.5 Further findings

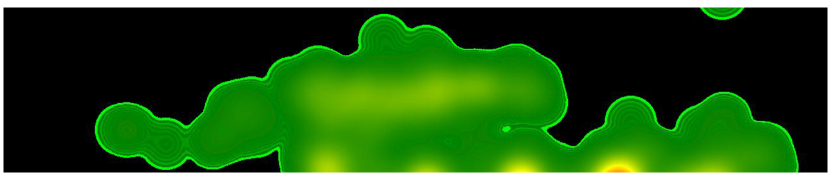
As data analysis reveals mean fixation duration has been significantly higher for true/false tasks compared to matching or search tasks on all tasks. Results indicate higher cognitive load and therefore more demanding efforts by participants. Longer mean fixation within eye movements go along with participants' feedbacks, as many volunteers have reported during retrospective analysis that had struggled with true/false tasks, because of its difficulty. One possible explanation for this outcome might be that task design for true/false needed more interpretation and understanding of the instructions provided in each task. For true/false tasks participants had not only to interpret the traffic signs' meanings, but also had to verify their this with the textual instruction. Concluding this, cognitive efforts to understand the instruction were higher compared to other task types. Heatmaps back up this assumption, as overall fixation intensities on the instruction sentences for true/false tasks were highest. Figure 44 shows that



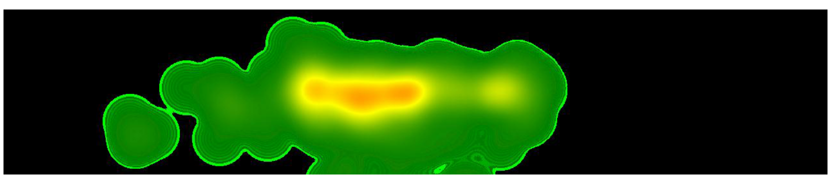
(a) Task instruction layout



(b) Overall heatmap of matching tasks' instructions



(c) Overall heatmap of search tasks' instructions



(d) Overall heatmap of true/false tasks' instructions

Fig. 44. Overall heatmap of true/false tasks' instructions had highest fixation intensities.

5.6 Discussion of qualitative results

Points to discuss so far: -Usability outcomes -During learning confirmatory scanning before final judgment have been recorded. -...

5.7 General recommendations

Only an idea, but maybe like this:

- When introducing unfamiliar and out-of-context visual elements cognitive load and intensity of visual percaption will increase.
- Eye movements within search tasks might be negatively correlated to matching or true/false tasks in terms of visual perception.
- Consider cultural differences, when designing online e-learning courses for students of different ethnic groups.
- If e-learning designers apply true/false tasks in their e-learning courses, they should be aware of that textual instructions have high visual importance, therefore attention should be laid on coherent phrasing.
- ...

6. STUDY LIMITATIONS

In this chapter factors are described that might had interferred the study's outcomes. Both limitations those of the design and also those (uncontrolable) factors of the date collection method are mentioned.

- **"Americanized" Chinese participants**: Due to financial restrictions the eye tracking study of the Chinese was conducted in the USA and not in China. For the study only those Chinese participants were accepted, who had residence not longer than 6 months. However as questionnaire reveals even the recruited volunteers showed higher knowledge on American traffic signs. Taking this aspect into account, it can be assumed that their eye movement on Americans signs is influenced in some way.
- **Chinese participants driving experience**: Chinese participants experience on driving ($M_{CN} = 3.7$ years) was significantly lower than those of the Austrian volunteers ($M_{AT} = 10.2$ years) and the American participants ($M_{US} = 8.3$ years). In order to this their level of knowledge might be different from the other ethnic groups.
- **Narrow group of participants**: As many publications on eye tracking (as an example see [Holmqvist et al. 2011]) critized that the majority of academic eye tracking research is carried out on university students distorting hereby eye tracking results due to an inappropriate sample. Taking this in account valid conclusions in this study can only be drawn for young academics aged between 20 and 35 years.
- **"Fake" traffic signs**: During the retrospective interview some participants have reported that during some task items they have believed that foreign signs are fake traffic signs and so, they have ignored them in their ongoing problem solving process. As a solution for future studies the participant's briefing must include a clarification that all traffic signs (visual objects in general) are real.
- **Number of participants**: For statistical analysis on ethnicity the number of participants for each ethnic group might be higher in order to have a larger sample. Nevertheless the authors believe that the results reveal tendencies and can serve as a sold basis for future work.
- **Chinese letters too small**: Two Chinese participants have reported that the Chinese letters and the spacing used in the e-learning environment made the teaching materials and the task instructions difficult and exhausting to read. To they end of the test session they have speeded up their pace. This effect was not evident in the pre-tests.
- **Lab scenario**: As the experiment was conducted in a laboratory and the outcom of the test had no further consequences for the participants, the e-learning tests did not entirely reflect a real test scenario.

7. CONCLUSION

In the present study the visual patterns of different ethnic groups on international traffic signs were recorded. The experiment was conducted in the context e-learning and the results yield some interesting effects. In contrast to other eye tracking studies investigating the ifluence of different goal and instruction, this study indicates teaching materials have no effect on tests about international traffic signs. Further research efforts should be given in evaluating the influence of teaching materials in other e-elarning topics or for participants having no driving experience. This study has also shown that sign origin and number of traffic signs in task items have strong significant effect for eye movements. The former provokes longer gaze duration, higher fixation counts and regression rates, whereas the latter has different effects depending on the task type. Other findings of the experiment show that there are some significant influences caused by culture. The most significant effect was revealed for Americans, who had lower mean fixation duration over the entire eye tracking sessions independently of task type or sign origin. More research is needed to clarify the reasons for this cultural difference. Furthermore the investigation yielded some interesting results in term of usability improvement of the Moodle e-learning environment. It is believed that the results might help to contribute to improve e-learning systems efficiency and to design more effective e-learning tasks. Moreover the results can contribute to develop more internationally valid traffic signs, which transmit more clearer messages for different cultures. Overall recogntion of traffic signs in driving education systems and so, higher awareness in driving can be achieved.

Acknowledgments

This work was partly supported by a grant of Vienna University of Technology for short-term academic work abroad.

REFERENCES

BARON, J. AND LI, Y. 2007. Notes on the use of R for psychology experiments and questionnaires. Online Notes. URL: `http://www.psych.upenn.edu/~baron/rpsych/rpsych.html` (last accessed Oct. 2012).

BENTE, G. 2005. Erfassung und analyse des blickverhaltens. *Lehrbuch der Medienpsychologie*, 297–324.

CHUA, H. F., BOLAND, J. E., AND NISBETT, R. E. 2005. Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences of the United States of America 102,* 35, 12629–12633.

EHMKE, C. AND WILSON, S. 2007. Identifying web usability problems from eye-tracking data. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it - Volume 1*. BCS-HCI '07. British Computer Society, Swinton, UK, 119–128.

FISHER, D. L., POLLATSEK, A. P., AND PRADHAN, A. 2006. Can novice drivers be trained to scan for information that will reduce their likelihood of a crash? *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention 12 Suppl 1*, i25–9.

GOLDBERG, J. H. AND KOTVAL, X. P. 1999. Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics 24,* 6, 631–645.

HOLMQVIST, K., NYSTRÖM, M., ANDERSSON, R., DEWHURST, R., AND JARODZKA, H. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.

INMAN, V. W. 2012. Conspicuity of traffic signs assessed by eye tracking and immediate recall. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 56. 2251–2255.

JACOB, R. J. K. AND KARN, K. S. 2003. Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. In *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, J. Hyönä, R. Radach, and H. Deubel, Eds. Elsevier Science, Amsterdam, The Netherlands, 573–605.

JUST, M. A. AND CARPENTER, P. A. 1976. Eye fixations and cognitive processes. *Cognitive Psychology 8*, 441–480.

KAAKINEN, J. K. AND HYÖNÄ, J. 2010. Task effects on eye movements during reading. *Journal of experimental psychology. Learning, memory, and cognition 36,* 6, 1561–6.

KO, M., HIGGINS, L., CHRYSLER, S. T., AND LORD, D. 2010. Effect of driving environment on drivers eye movements: Re-analyzing previously collected eye-tracker data. In *Transportation Research Board 89th Annual Meeting*. Washington, DC, USA.

NG, A. W. Y. AND CHAN, A. H. S. 2007. Cognitive design features on traffic signs. *Engineering Letters 14*, 13–18.

NIELSEN, J. AND PERNICE, K. 2010. *Eyetracking Web Usability*. Voices That Matter. New Riders, Berkeley, CA.

PALINKO, O., KUN, A. L., SHYROKOV, A., AND HEEMAN, P. 2010. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ETRA '10. ACM, New York, NY, USA, 141–144.

PEETERS, M. 2010. Consumers information needs on e-commerce websites: A cross-cultural eye tracking study.

PERNICE, K. AND NIELSEN, J. 2009. Eyetracking methodology – how to conduct and evaluate usability studies using eye tracking. Tech. rep., Nielsen Norman Group, Fremont, CA, USA. August.

RAKOCZI, G. 2012. Eye tracking in forschung und lehre: Möglichkeiten und grenzen eines vielversprechenden erkenntnismittels. *Digitale Medien: Werkzeuge für exzellente Forschung und Lehre*, 87–98.

SALVUCCI, D. D. AND GOLDBERG, J. H. 2000. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Eye Tracking Research & Applications (ETRA) Symposium*. ACM, Palm Beach Gardens, FL, 71–78.

TAYA, S., WINDRIDGE, D., AND OSMAN, M. 2012. Looking to score: the dissociation of goal influence on eye movement and meta-attentional allocation in a complex dynamic natural scene. *PloS one 7,* 6.

YAN, H., GUAN, H., AND XU, Z. *Incorporating Saliency Map into Prediction of Drivers' Eye Fixations on Traffic Signs*. Chapter 161, 1627–1639.

YEN, N.-S., TSAI, J.-L., CHEN, P.-L., LIN, H.-Y., AND CHEN, A. L. P. 2011. Effects of typographic variables on eye-movement measures in reading chinese from a screen. *Behaviour & IT 30,* 6, 797 – 808.

ZHU, Z. AND JI, Q. 2004. Real time and non-intrusive driver fatigue monitoring. In *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*. 657–662.

Received 2; revised 0; accepted 1

2