Who can benefit from augmented reality in chemistry? Sex differences in solving stereochemistry problems using augmented reality

Sebastian Habig

Sebastian Habig is a chemistry education researcher at the University of Duisburg-Essen in Germany. In his research, he is interested in questions regarding the use of ICT in chemistry education and how digital media can be designed and implemented to support students’ chemistry learning. Address for correspondence: Dr Sebastian Habig, University of Duisburg-Essen, Schützenbahn 70, Essen 45127, Germany. Email: sebastian.habig@uni-due.de

Abstract

This paper is about augmented reality (AR) and its potentials to support students in handling scientific representations. For this purpose, first representations are examined from a science educational and instructional psychology perspective. After giving a short overview of AR in general and how it can be delineated from virtual reality (VR), potential advantages of an educational use of AR are outlined considering typical difficulties of students when learning with scientific representations. Since literature frequently reports sex differences in spatial abilities, this study focuses on potential differential effects in the use of AR depending on sex. Against this background, AR might be a tool which can help to compensate for disadvantages in spatial abilities. In order to investigate this question, chemistry students had to answer 20 items related to stereochemistry, a concept that places high demands on mental spatial rotation skills. While one half of the items had to be solved using two-dimensional (2D) ball-and-stick figures of molecules, the other half had to be answered with the help of AR representations. Due to the aforementioned sex differences regarding visuo-spatial abilities between males and females found in the literature, it is hypothesized that AR representations will support all students but females in particular by reducing cognitive load. If this assumption is correct, the AR items would have to be correctly solved by the students more often than the 2D items. The results of analyses of variance indeed reveal a significant effect of the sex variable dependent on the type of representation. In addition, a questionnaire was administered to survey the students’ attitudes towards learning with the AR app used.

Introduction

In all STEM domains (science, technology, engineering and mathematics) different kinds of representations are used to illustrate scientific concepts with each domain having specific rules and conventions for using and interpreting representations. Especially in chemistry, there is a huge variety of representations and their use is associated with different objectives. For example, chemists use wedge and dash projections of molecular structures to visualize chemical binding properties between atoms or, another example, atomic and molecular orbitals to illustrate the probability density of finding an atom’s electron in any specific region around the nucleus both
mathematically and graphically. In addition, chemical symbols and chemical equations can be seen as chemistry-specific representations which are necessary for the development of conceptual scientific knowledge and, moreover, to communicate in a scientific community (Coleman, McTigue, & Smolkin, 2011; Hoffmann & Laszlo, 1991; Oliveira et al., 2013). Due to the importance of representations in scientific contexts it should not be presupposed that students automatically know how to interpret representations and how they can contribute to their learning (Dickmann, Opfermann, Dammann, Lang, & Rumann, 2019; Gilbert, 2008; Rau, 2017). For many students, it is difficult to extract information from representations that is relevant for their learning process (Oliver-Hoyo & Sloan, 2014). Translating external representations like 2D visualizations into three-dimensional (3D) mental models, eg, is a highly challenging cognitive process that can easily lead to cognitive overload, especially when the learning content itself is highly sophisticated (Wu & Shah, 2004). Because of previous research that reports sex differences with regard to spatial ability, these difficulties specifically apply to girls and women respectively (Castro-Alonso & Jansen, 2019; Harle & Towns, 2011; Kimura, 2000; Terlecki & Newcombe, 2005).

A promising approach for supporting students to meaningfully integrate chemistry-related external representations into their learning process is to integrate augmented reality (AR) models into conventional text-based instructional material (Chen, Teng, & Lee, 2011; Cheng & Tsai, 2013; Musio et al., 2016). Against the background of multimedia learning theories (Mayer, 2014; Schnozt & Bannert, 2003), it can be assumed that AR visualizations may support learners in selecting and organizing relevant spatial information and hence lead to higher learning outcomes. Although the employment of AR in educational contexts has received widespread academic attention, only little has yet been said about the specific conditions and student characteristics that determine its effectiveness (Ibáñez & Delgado-Kloos, 2018).

This paper ties in with this initial situation. First, from an instructional psychology and chemistry education point of view, it is illustrated what is meant when the term multiple external
representations is discussed and what role it plays for learning. After taking a closer look at sex differences regarding spatial abilities, insights into the effectiveness of AR in educational contexts, especially for chemistry, will be summarized. Subsequently, on the basis of the theoretical framework, research questions are derived that were examined within this study. After describing the study design, the sample and the test instruments used, the results are presented. Finally, the results are discussed against the background of previous findings and limitations of the study are pointed out.

**Representations in science education and instructional psychology**

Dealing with scientific representations is part of the daily work of professionals in chemistry and it is also a key component of chemistry studies at both the school and university level. The competence to cognitively process multiple external representations is crucial for the learning of new contents and concepts (de Vries, Demetriadiis, & Ainsworth, 2009; National Research Council, & Geographical Sciences Committee, 2006; Opfermann, Schmeck, & Fischer, 2017; Wu & Krajcik, 2006; Yang, Andre, Greenbowe, & Tibell, 2003). Representations which are embedded in instructional material like textbooks are called external representations. Wu and Puntambekar (2012) differentiate the following forms of external representations.

- **Verbal–textual**: This type of representation covers all kinds of verbal- and text-based information.
- **Symbolic–mathematical**: This type of representation includes chemical and mathematical equations and typical chemical visualizations like structural formula or ball-and-stick models.
- **Visual–graphical**: In contrast to symbolic–mathematical representations visual–graphical representations refer to real-life phenomena and thus the macroscopic level. Usually, animations and simulations are examples for visual–graphical representations.
- **Actional–operational**: This type of representation includes hands-on experiments and inquiry-based actions.

The four mentioned types of external representations are not independent of one another which makes a clear distinction difficult. Computer simulated experiments (visual–graphical), eg, can also be enhanced by symbolic representations which illustrate processes on the microscopic or sub-microscopic level. Furthermore and because of the rapid technological development, other innovative representation types are conceivable which are not listed above (Wu & Puntambekar, 2012).

While few studies contrasted the effects of learning with single external representations the vast majority of studies in this context investigated the impact of different combinations of multiple external representations on learning (Dickmann et al., 2019; Gilbert, 2008; Rau, 2017; Treagust & Tsui, 2013; Wu & Puntambekar, 2012). For example, Treagust, Chittleborough, and Mamiala (2003) point out that using multiple external representations simultaneously can support students in making connections between different concepts resulting in deeper understanding.

From an instructional psychology perspective positive effects of multiple external representations can be explained against the background of Mayer’s Cognitive Theory of Multimedia Learning (CTML, Mayer, 2014) and the Integrated Model of Text and Picture Comprehension (IMTPC, Schnotz & Bannert, 2003). Both theories act on the assumption that the working memory’s capacity is limited. In addition, based on Paivio’s dual-coding theory it is assumed that verbal and visual information are processed in two distinct cognitive channels: the auditory/verbal and the visual/pictorial channel. Dependent on the characteristics of the presented information it is processed either in the auditory/verbal or in the visual/pictorial channel of the working memory. On
the one hand, following the IMTPC, the verbal information processing results in a propositional network which integrates essential information of the external representation. On the other hand, the visual information processing leads to an internal mental model of the visual representation. The successful alignment of both propositional network and internal mental model including prior knowledge stored in the long-term memory can result in meaningful learning. However, the information processing of novices in a learning domain is often impaired due to high cognitive load which can be caused by contents of high complexity or poorly designed learning material.

Learners need well developed representation competences in order to generate adequate internal mental models from (multiple) external resources. As stated above, it cannot be presupposed that learners bring these competences, especially when they are novices in a learning domain (NRC, 2006; Rau, 2017). Commonly learners are expected to learn with external, domain-specific representations although they might not be able to integrate them into their learning process, properly. They have to know how to interpret external representations and how they can select and organize relevant information from them. In the literature this conflict is referred to as “representation dilemma” (Rau, 2017). This problem intensifies when effects of general spatial ability are considered. Correlational studies show that it is more difficult for learners with low spatial abilities to solve science-related problems even though the specific tasks do not require spatial skills to be solved (Wu & Shah, 2004). Students with low spatial skills often fail at translating external representations into internal mental models. This can be explained by taking a closer look at the cognitive processes which are involved when a 2D visualization has to be translated into a 3D mental model. This process is accompanied by high cognitive demands which can easily overwhelm students with low spatial skills. Thus, without supporting this process by appropriate instructional methods cognitive overload can occur, which specifically applies to novices in a learning domain (Urahne, Nick, & Schanze, 2009; Wu & Shah, 2004). Accordingly, using (multiple) external representations might help those students who already have high spatial skills and can process the spatial information adequately, in particular (Gyselinck, Cornoldi, Dubois, Beni, & Ehrlich, 2002; Opfermann, Schmeck, & Fischer, 2017).

Against the background of the relation between spatial ability and learning performance, spatial ability research literature shows that girls and women are particularly disadvantaged, because of sex differences in spatial abilities (Castro-Alonso & Jansen, 2019; Halpern & Collaer, 2005; Kimura, 2000; Terlecki & Newcombe, 2005). In their review of spatial ability literature Harle and Towns (2011, p. 354) point out that “the largest differences are found in tasks involving 3D rotation that show effect sizes of nearly 1.0 standard deviation.” Reasons for the origin of these differences are divers and range from physiological to developmental arguments. However, chemistry educators and chemistry education researchers should be aware of these differences when they plan their teaching or learning material. Especially in organic chemistry 3D rotation is omnipresent and crucial for the understanding of molecular geometry and reaction mechanisms. Fortunately, many studies in this context suggest that spatial skills can be improved with appropriate interventions and trainings (Barnea & Dori, 2002; Castro-Alonso & Uttal, 2019; Harle & Towns, 2011; Wu & Shah, 2004).

The question arises, how instructional techniques can support students in translating external representations into internal mental models. Wu and Shah (2004) describe five design principles which should be considered for the development of learning material which take the aforementioned problems into account.

1. Multiple rather than single external representations should be included.
2. Relations between multiple external representations should be highlighted explicitly.
3. Dynamic and interactive processes should be made visible.
4. The translation from 2D into 3D representations should be supported actively.
5. Cognitive load should be reduced by presenting relevant information near to each other (contiguity principle, Mayer, 2014).

A promising approach which can cover all of the five listed design principles is embedding augmented reality (AR) technology into learning environments and material. A definition of AR and a brief overview of existing research on the use of AR in science education contexts is given in the following section.

Augmented reality in educational contexts

The term augmented reality describes a technology that allows to overlay virtual, computer-based elements into real world environments in real time (Azuma, 1997). The rapid technological development makes AR technology available for everyone by operating on smartphones or tablets, for instance. With the help of AR, digital elements like texts, directional signs, 3D models or animations can directly be projected on real-world objects.

Studies which investigated the effects of AR applications in different fields of education show that, in general, AR can foster cognitive and affective learning outcomes of students (Akçayır & Akçayır, 2017; Bacca, Baldiris, Fabregat, & Graf, 2014; Cai, Liu, Yang, & Liang, 2019; Cheng & Tsai, 2013). For STEM education, Ibáñez and Delgado-Kloos (2018) show that the cognitive outcome which is evaluated in the majority of quantitative studies is knowledge acquisition. For low-level cognitive processes like retention over brief periods of time, positive effects of AR supported learning could already be found. However, there are few empirical studies which take higher order cognitive skills like critical thinking, problem-solving or spatial reasoning into account (Jaramillo, 2017). Another encouraging finding of AR in STEM educational contexts is that most studies report positive effects on affective variables of students like motivation and attitudes towards STEM subjects. Against the background of decreasing interest of students in STEM subjects, this may help to design appealing instructional material with which students enjoy learning (Cheung, 2017; Yu, Sun, & Chen, 2017; Zimmermann, Land, & Jung, 2016).

A main reason which distinguishes AR technology from other instructional techniques is that it can help to visualize phenomena (e.g., magnetic field lines) that otherwise could not be observed (Klopfer & Squire, 2008). Nielsen and colleagues provide an overview of AR approaches in STEM subjects (Nielsen, Brandt, & Swensen, 2016). The authors introduce an AR application which enables students to “look” into the bodies of their classmates by using a 3D model of the human respiratory system. The researchers mention that it was especially advantageous for students to have realistic, 3D and animated visualizations which illustrate complex learning contents.

Another example for how AR can be used in STEM educational contexts is to combine interactive AR elements and conventional text-based instruction (Chen, Teng, & Lee, 2011; Cheng & Tsai, 2013; Musio et al., 2016). So-called marker-based AR applications use predefined markers (e.g., images, signs, etc) to anchor the location where a 3D model should be projected. With the help of a device’s camera function, the AR application is able to recognize the corresponding marker and overlays the predefined and interactable 3D model on its location (El Sayed, Zayed, & Sharawy, 2011; Pence, 2010). Figure 1 shows three marker-based AR representations, developed by the author.

AR markers can easily be combined with print media like textbooks or worksheets and help to visualize 3D objects close to textual information which is beneficial in terms of learning with multimedia (contiguity principle, Mayer, 2014).
Particular potentials for the use of AR are also seen, especially for the field of chemistry. Zheng and Waller (2017) emphasize that AR can be used to create and modify molecular structures using hand gestures. In addition, they emphasize the interactivity of AR visualizations as an advantage, since, eg, binding lengths or angles can also be measured using hand gestures (Zheng & Waller, 2017). The interactivity of AR learning settings is also the focus of a study by Cai, Wang, and Chiang (2014). The authors use an interactive learning setting in which students learn about atomic structure and chemical bonds with the help of AR models. The authors assume that AR can help students to understand the connection between the phenomenological and the sub-microscopic level. They report positive effects of the AR learning settings on learning performance and attitudes compared to a comparison group (Cai, Wang, & Chiang, 2014). One weakness of the study, however, is that it lacks a critical examination of possible student misconceptions. The question remains as to which misconceptions might be promoted by the use of the AR application, since the students might gain the impression that properties of a substance can be transferred directly to the sub-microscopic level.

Berson, Ng, Shin, and Jenkinson (2018) compared in an experimental study how students’ understanding of stereochemical concepts changes when they learn with AR models, computer-based on-screen models or physical models. One result of the study is that the students who learned with the AR models, under control of prior knowledge, achieved higher test scores in a corresponding posttest than the students of the other groups (Berson et al., 2018). However, individual differences in spatial abilities of the subjects were not considered. This reveals a research gap, as many studies report positive effects of AR learning opportunities on cognitive factors such as student learning performance, affective factors like interest and perceived self-efficacy (Berson et al., 2018; Cai, Wang, & Chiang, 2014; Nechypurenko, Starova, Selivanova, Tomilina, & Uchitel, 2018; Qassem, Hawai, Shehhi, Zemerly, & Ng, 2016; Wan, San, & Omar, 2018; Zheng & Waller, 2017). However, the role of spatial abilities and related sex differences in the work with AR in chemical contexts has so far been poorly investigated.

Aim of the study and research questions
The main aim of the study is to investigate whether students of a bachelor chemistry programme are able to use AR representations to solve domain-specific tasks. Therefore, it will be examined
whether students perform better on a test on stereochemistry when they are provided with AR representations instead of 2D visualizations. In the light of sex differences regarding spatial ability mentioned in the research literature it will further be investigated if females benefit from AR representations to a higher degree compared to males. In addition, students’ attitudes towards learning with AR representations will be assessed.

This study addresses the following research questions:

RQ1: Do students determine the absolute configuration of chiral molecules more often correct when AR representations are provided?

RQ2: Can sex differences with regard to RQ 1 be identified?

RQ3: How do students evaluate the learning potentials of the AR App and their interest in learning with it in general?

By providing 3D spatial information through AR, it is assumed that students solve items more often correct which are based on this specific form of visualization (H1). Further it is expected that females will benefit most from AR representation because this form of visualization might contribute to bridge the apparent gap regarding mental rotation ability (H2).

Research question three is rather explorative which allows several assumptions. On the one hand, it can be assumed that technology-oriented students in particular prefer learning with AR representations. On the other hand, it could be possible that the students criticize the usability of the used web application or are sceptic about new technologies in general and therefore reject to integrate them into their learning.

Design and methods

In order to answer research questions one and two, we developed a test on stereochemistry in organic chemistry with 20 items. While most of the items require to determine the absolute configuration of an organic structure, four items are about structures’ conformation. A total of 20 items were included in the initial version of the test so that a sufficient range of item difficulties could be realized. In the initial version of the test, there are a total of four items regarding determining the conformation of chemical structures. The other items deal with the determination of the absolute configuration of different molecules, whereby the number of stereo centres which have to be determined varies. A further four items were recorded in which there was only one stereo centre to determine (slightly sophisticated). For another 10 items, there were two stereo centres per structure to be determined (medium sophisticated) and, for another two structures, three stereo centres (highly sophisticated) had to be named.

Each task consists of a ball-and-stick representation of a specific chemical structure which is needed to solve the task. Half of the items are based on 2D figures and AR representations respectively. All items are designed in a multiple-choice single-select format. It would also have been possible to have an open answer format which would have required students to identify the entire structures on their own. However, since the tasks only focused on the correct naming of the stereo centres and not on naming the entire structure, the multiple-choice format seems appropriate, especially as suitable distractors could be created relatively easily for this type of task. Figure 2 shows examples for both item types.

In case of the AR item type the ball-and-stick figure is replaced by a so-called AR marker. The participants visualize the underlying AR representation by opening the app ARC—Augmented Reality Chemistry and scanning the marker. The app was developed with the software Unity 3D and the Vuforia Engine. The 3D models used could be exported from the PubChem database using
the Jmol software and, after converting, imported into the Unity 3D environment. For this study, a specific user interface was not used so that the students had direct access to the AR camera when starting the app. Each participant was provided with an Apple iPad (9.7", 6th Gen.) with the pre-installed app. In total, the version of the app used for this study included the ten 3D AR models (ten markers) needed to complete the test. Once the AR model is visible it can be rotated and rescaled by common gesture controls. The test was administered within a regular organic chemistry course. Figure 3 illustrates how the app is used.

Figure 2: Item examples

Figure 3: Visualizing AR models using the developed web app
Sample and instruments
The study was conducted in summer 2018 and 31 (16 female, 15 male) students participated in our study. They all were enrolled in a bachelor chemistry programme at the University of Duisburg-Essen in Germany. In addition to the stereochemistry test, the participants worked on a short version of the Purdue Visualization of Rotation Test (Bodner & Guay, 1997). We administered this test in order to gain a proxy for the students’ mental rotation ability. The test is about reconstructing a given rotation of an object (a), transferring this rotation to a new object (b) and finally, selecting the correct result from a given choice (c). The tested underlying cognitive processes are comparable to what the students have to perform in the stereochemistry test. The original test consists of 30 items and was evaluated in many studies with appropriate large sample sizes (Bodner & Guay, 1997). However, due to time restrictions, we decided to use a shorter version of the test consisting of eight items.

In addition to both tests and in light of research question three, we additionally asked the students to fill in a questionnaire and rate their experience with the AR application. Furthermore, we were interested in the students’ opinions about learning with AR. The items had to be rated on a 4-point Likert-type scale ranging from 1—strongly agree to 4—strongly disagree. Table 1 provides an overview of the items, which had to be rated by the students (translated from German).

To answer the stereochemistry test the students needed prior knowledge regarding the determination of the absolute configuration of chiral molecules with the help of the Cahn–Ingold–Prelog (CIP) nomenclature. To ensure this, it was required that the students already passed the basic course in organic chemistry before they were allowed to participate in our study. Nevertheless, we provided basic CIP rules and the priority of the specific substituents in each items’ description (See Figure 2).

Table 1: Statements regarding the use of AR representations, which had to be rated by the participating students (translated from German)

| Please rate the following statements: | Strongly agree | Agree | Disagree | Strongly disagree |
|--------------------------------------|----------------|-------|----------|------------------|
| I can imagine to learn organic chemistry with the help of AR representations. | ☐              | ☐     | ☐        | ☐                |
| AR representations are a reasonable extension to 2D visualizations. | ☐              | ☐     | ☐        | ☐                |
| AR representations can help me to understand 3D information. | ☐              | ☐     | ☐        | ☐                |
| I think Organic Chemistry textbooks with integrated AR representations can be helpful. | ☐              | ☐     | ☐        | ☐                |
| For me it is difficult to imagine things in 3D space. | ☐              | ☐     | ☐        | ☐                |
| Basically, I assume that AR representations may help to present spatial information. | ☐              | ☐     | ☐        | ☐                |
| I think it would be interesting to learn with AR representations. | ☐              | ☐     | ☐        | ☐                |
| Using AR representations was fun. | ☐              | ☐     | ☐        | ☐                |
| The tasks with AR support were easier. | ☐              | ☐     | ☐        | ☐                |
| It was too complicated to use the AR representations. | ☐              | ☐     | ☐        | ☐                |
| Usually, I use digital media for learning (eg, instructional videos, simulations, etc). | ☐              | ☐     | ☐        | ☐                |
Results

Reliability analysis
In a first step, we analysed the developed test on stereochemistry regarding its internal consistency. Therefore, we conducted a reliability analysis with all 20 items resulting in a nonsatisfying Cronbach’s Alpha measure of $\alpha = .39$. Further analyses on the item level revealed negative item-total correlations for six test items. Taking a deeper look at the questionable items, it became apparent that those items showed bad fit statistics which were based on chemical structures with five different substituents or three stereo centres respectively. It can be assumed that these items were too difficult for the participating students, resulting in a poor item-total correlation. A reapplied reliability analysis without the respective items resulted in an acceptable internal consistency measure ($\alpha = .70$). Therefore, for following analyses the remaining 14 test items are considered. The final test can be downloaded here http://udue.de/arsctest.

Comparison of the item types
Based on the final 14 items a test score (sum score) was generated for each participant. Due to a dichotomous item coding (correct/incorrect) the score ranged from 0 to 14 points. Since 7 items were based on 2D visualizations and a further 7 items on AR representations, a maximum of 7 points could be achieved for each test part. The combined result of both scales is the average overall test result of the students $M = 6.63$ ($SD = 2.82$).

In order to compare the students’ test performance for both item types, item-specific mean scores were calculated. Figure 4 shows the mean sum scores for the two item types (2D: $M = 3.25$, $SD = 1.59$ and AR: $M = 3.40$, $SD = 1.67$) separately. To give an answer to research question one a paired t-test was conducted to compare the mean scores. Initially, the result shows no significant difference of both test scores ($t(29) = 0.542$, $p = .59$).

![Figure 4: Mean sum scores for both item types](image-url)
Against the background of possible sex differences in working with different kinds of representations and to address the second research question an analysis of variance (ANOVA) was conducted. While both test score variables were considered as dependent variables in the model, the dichotomous sex variable (male/female) was included as single factor. The analysis reveals a significant effect of the sex variable on test performance \( (F(1, 28) = 6.375, p = .018, d = .95) \). Figure 5 shows that, on the one hand, females score higher on items that are based on 2D figures, whereas on the other hand, males solve AR-based items more often correct. Further post hoc analyses of these findings show that while the scores between males and females do not differ significantly in the 2D part of the test \( (F(1, 28) = .134, p = \text{n.s.}) \), there is a significant difference in the AR part of the test \( (F(1, 28) = 7.428, p = .011, d = 1.03) \).

In the light of sex differences regarding spatial ability found in the research literature, a confounding effect of this construct may be assumed. Therefore, an additional ANOVA was conducted, including the students’ individual test score of the Purdue Visualization of Rotation Test as a covariate. The results show a significant effect of the covariate on the dependent variables \( (F(1, 28) = 5.073, p = .033, d = .87) \). However, the effect of the sex variable does not only remain significant, it even increases \( (F(1, 28) = 8.497, p = .007, d = 1.12) \).

Students’ attitudes towards learning with AR representations

The following exploratory part of the results section will give a first view on how students believe to use AR representations for their learning. Figure 6 shows mean scores for the items listed in Table 1 separated for males and females (1—strongly disagree; 2—disagree; 3—agree; 4—strongly agree).

Tendentially, the participating students see a potential benefit of AR representations and view them as meaningful supplement for 2D visualizations (Item 2: \( M = 3.64, SD = .49 \)). In addition, they report high interest in learning with AR visualizations (Item 7: \( M = 3.32, SD = .72 \)) and they enjoyed working with them (Item 8: \( M = 3.14, SD = .76 \)). It has to be mentioned that no significant differences between males and females regarding the 11 administered items were found.
Furthermore, no significant effect of one of the variables can be found when they are included in
the previously conducted ANOVA model.

Discussion and implications
The study’s first research question is: Do students determine the absolute configuration of chiral molecules more often correct when AR representations are provided? It was assumed that students solve items more often correct which are based on AR representations by providing 3D spatial information (H1). Without any further restrictions the question and related hypothesis have to be rejected. However, the presented results show that male participants solved items more frequently correct when they were based on AR representations rather than conventional 2D figures. In contrast, it seems that females were not able to use the AR representations to correctly solve the items. They instead scored higher on items based on 2D ball-and-stick models of molecules. Therefore, H1 can only be confirmed for male students.

The second research question extends RQ1 by taking possible sex differences into account. Contrary to the initial assumption that AR representations should be most effective for females due to a compensating effect regarding disadvantages in spatial abilities (H2), the results show that males in particular benefit from this kind of representation. At a first glance, the results of this study therefore contradict findings in the literature. Since the tasks of the stereochemistry test mainly require mental rotation ability, sex effects would have been expected for both parts of the test, as reported eg, by Castro-Alonso and Jansen (2019) and Harle and Towns (2011).

© 2019 The Authors. British Journal of Educational Technology published by John Wiley & Sons Ltd on behalf of British Educational Research Association
Looking only at the results of the students in the 2D part of the stereochemistry test, there are no statistically significant differences. Females solve these tasks as often as men. Interestingly, males perform significantly better than females in the AR items. A possible reason for this could be that males may be more able to use AR as an instructional tool to solve the tasks of the test. It should be highlighted that the found effect does not diminish or even disappear when mental rotation ability is controlled. Contrary to the initial expectation the effect even increases. Therefore, the assumed compensating effect of AR representations cannot be approved in this study.

As an alternative it can also be argued that the findings of this study are in line with the results of a meta-analysis by Castro-Alonso, Wong, Adesope, Ayres, and Paas (2019), which showed that studies on the efficacy of dynamic visualizations report larger effect sizes when there were disproportionately large numbers of males in the corresponding sample. Since the AR representations used in this study were interactive and could be manipulated by the users, it can be assumed against the background of the literature that this is of particular benefit to the male participants thus resulting in the found sex difference regarding the AR part of the test. The authors argue that the sex variable should increasingly be considered in research on instructional visualizations which is also supported by the results of this study (Castro-Alonso et al., 2019).

Similar results are also reported by Berson et al. (2018). The authors summarize that learners who learned stereochemistry with AR representations achieved higher test scores in a posttest than students who learned with on-screen or physical representations. Although this study by Berson et al. (2018) is a comparison of learning and not test situations, parallels to this study can be found. Unfortunately, the sex of the participants was not considered, so that no further comparisons can be made here.

The findings of this study raise new questions about the relation between AR representations, spatial abilities and sex. Why seem males to be more likely able to use AR representations to correctly solve specific items on stereochemistry? Is it easier for them to extract relevant 3D information from such representations? A possible explanation may be that the participants had to combine textual (priorities of the substituents) and visual information in the course of the test. The required cognitive processes might be hard to perform for students with low spatial abilities leading to high cognitive load (van Bruggen, Kirschner, & Jochems, 2002). In order to reduce cognitive load students might be trained in using AR representations in a first step. According to that, further research should focus the question how AR representations can best be implemented into instructional material and if design principles can be identified (Ibáñez & Delgado-Kloos, 2018). Especially in context of designing appropriate learning material it would be interesting to learn more about how students integrate AR representations into their information processing and how sex differences can be explained.

For science education in general and chemistry education in particular AR technology provides a promising approach to visualize complex concepts. This is supported by the consistently positive ratings of students regarding learning with AR representations. There is a need for further studies on the effectiveness of learning chemistry with AR which can help to formulate appropriate design principles.

Limitations
A clear limitation of the study is the relatively small sample size of $N = 31$ students. Since the study was conducted within a regular Organic Chemistry course the number of participants was limited. Therefore, the reported findings need to be replicated with a larger sample size. In addition, it would make sense to test the usability of the app used within the framework of a usability study. Another limitation of the study is the lack of a distinct measure of domain-specific prior
knowledge. Although the students were only allowed to participate when they already passed the basic course in organic chemistry, they differ in respect of prior knowledge which should be controlled for. However, in this study, all important additional information like the substituents’ priorities were provided in order to reduce the effect of prior knowledge. Finally, it would be thinkable to include other chemical contents and concepts which put high requirements on students’ spatial abilities.

Acknowledgement
I thank Merve Ceylan, Alper Öztürk and Christoph Pelka (all University of Duisburg-Essen) for assistance with the data collection and app development. I would also like to thank the reviewers whose comments have greatly improved the manuscript.

Statements of open data, ethics and conflict of interest
The dataset generated and analysed during the current study are not publicly available due to privacy reasons but are available from the author on reasonable request.

The research adhered to ethical standards and guidelines as the nature of study demanded. The statistical analysis was performed using non-identifiable data.

The author has no conflicts of interest to declare in relation to this work.

References
Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. Educational Research Review, 20, 1–11. https://doi.org/10.1016/j.edurev.2016.11.002

Azuma, R. T. (1997). A survey of augmented reality. Presence: Teleoperators & Virtual Environments, 6(4), 355–385.

Bacca, J., Baldiris, S., Fabregat, R., & Graf, S. (2014). Augmented reality trends in education: A systematic review of research and applications. Journal of Educational Technology & Society, 17(4), 133.

Barnea, N., & Dori, Y. J. (2002). High-school chemistry students’ performance and gender differences in a computerized molecular modeling learning environment. Journal of Science Education and Technology, 8(4), 257–271.

Berson, M., Ng, D., Shin, J., & Jenkinson, J. (2018). Assessing augmented reality in helping undergraduate students to integrate 2D and 3D representations of stereochemistry. Journal of Biocommunication, 42(1). Retrieved from https://journals.uic.edu/ojs/index.php/jbc/article/download/9187/7547

Bodner, G. M., & Guay, R. B. (1997). The Purdue visualization of rotations test. The Chemical Educator, 2(4), 1–17.

Cai, S., Liu, E., Yang, Y., & Liang, J.-C. (2019). Tablet-based AR technology: Impacts on students’ conceptions and approaches to learning mathematics according to their self-efficacy. British Journal of Educational Technology, 50(1), 248–263. https://doi.org/10.1111/bjet.12718

Cai, S., Wang, X., & Chiang, F.-K. (2014). A case study of augmented reality simulation system application in a chemistry course. Computers in Human Behavior, 37, 31–40. https://doi.org/10.1016/j.chb.2014.04.018

Castro-Alonso, J. C., & Jansen, P. (2019). Sex differences in visuospatial processing. In J. C. Castro-Alonso (Ed.), Visuospatial processing for education in health and natural sciences (Vol. 29, pp. 81–110). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-20969-8_4

Castro-Alonso, J. C., & Uttal, D. H. (2019). Science education and visuospatial processing. In J. C. Castro-Alonso (Ed.), Visuospatial processing for education in health and natural sciences (Vol. 13, pp. 53–79). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-20969-8_3

© 2019 The Authors. British Journal of Educational Technology published by John Wiley & Sons Ltd on behalf of British Educational Research Association
Who can benefit from augmented reality in chemistry?

Castro-Alonso, J. C., Wong, M., Adesope, O. O., Ayres, P., & Paas, E. (2019). Gender imbalance in instructional dynamic versus static visualizations: A meta-analysis. Educational Psychology Review, 27(3), 1. https://doi.org/10.1007/s10648-019-09469-1

Chen, N.-S., Teng, D. C.-E., & Lee, C.-H. (2011). Augmenting paper-based reading activity with direct access to digital materials and scaffolding questions. Computers & Education, 57(2), 1705–1715.

Cheng, K.-H., & Tsai, C.-C. (2013). Affordances of augmented reality in science learning. Suggestions for future research. Journal of Science Education and Technology, 22(4), 449–462.

Cheung, D. (2017). The key factors affecting students’ individual interest in school science lessons. International Journal of Science Education, 30(4), 1–23.

Coleman, J. M., McTighe, E. M., & Smolkin, L. B. (2011). Elementary teachers’ use of graphical representations in science teaching. Journal of Science Teacher Education, 22(7), 613–643.

de Vries, E., Demetriadis, S., & Ainsworth, S. (2009). External representations for learning. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), Technology-enhanced learning (pp. 137–153). Dordrecht: Springer.

Dickmann, T., Opfermann, M., Dammann, E., Lang, M., & Rumann, S. (2019). What you see is what you learn? The role of visual model comprehension for academic success in chemistry. Chemistry Education Research and Practice. https://doi.org/10.1039/C9RP00016J

El Sayed, N. A. M., Zayed, H. H., & Sharawy, M. I. (2011). ARSC: Augmented reality student card. Computers & Education, 56(4), 1045–1061.

Gilbert, J. K. (2008). Visualization: An emergent field of practice and enquiry in science education. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Eds.), Visualization: Theory and practice in science education (pp. 3–24). Dordrecht: Springer.

Gyselinck, V., Cornoldi, C., Dubois, V., de Beni, R., & Ehrlich, M.-F. (2002). Visuospatial memory and phonological loop in learning from multimedia. Applied Cognitive Psychology, 16(6), 665–685.

Halpern, D. F., & Collaer, M. L. (2005). Sex differences in visuospatial abilities. In P. Shah, & A. Miyake (Eds.), The Cambridge handbook of visuospatial thinking (pp. 170–212). Cambridge University Press.

Harle, M., & Towns, M. A. (2011). Review of spatial ability literature, its connection to chemistry, and implications for instruction. Journal of Chemical Education, 88(3), 351–360.

Hoffmann, R., & Laszló, R. (1991). Representation in chemistry. Angewandte Chemie, 30, 1–16.

Ibáñez, M.-B., & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. Computers & Education, 123, 109–123.

Jaramillo, S. G. (2017). Horizon Report-2017 Higher Education Edition. Cuaderno Activa. 9(9), 171.

Kimura, D. (2000). Sex and cognition. Cambridge: MIT Press.

Klopfer, E., & Squire, K. (2008). Environmental detectives—The development of an augmented reality platform for environmental simulations. Educational Technology Research and Development, 56(2), 203–228.

Mayer, R. E. (2014). The Cambridge handbook of multimedia learning. Cambridge: Cambridge University Press.

Musio, B., Mariani, F., Śliwiński, E., Kabeshov, M., Odajima, H., & Ley, S. (2016). Combination of enabling technologies to improve and describe the stereoselectivity of wolff–staudinger cascade reaction. Synthesis, 48(20), 3515–3526.

National Research Council, & Geographical Sciences Committee. (2006). Learning to think spatially. Washington, D.C: National Academies Press.

Nechypurenko, P. P., Starova, T. V., Selivanova, T. V., Tomilina, A. O., & Uchitel, A. D. (2018, October 2). Use of augmented reality in chemistry education. In Proceedings of the 1st International Workshop on Augmented Reality in Education (pp. 15–23). Kryvyi Rih, Ukraine. (2257). ISSN 1613–0073.

Nielsen, B. L., Brandt, H., & Swensen, H. (2016). Augmented Reality in science education–affordances for student learning. Nordic Studies in Science Education, 12(2), 157–174.

Oliveira, A. W., Rivera, S., Glass, R., Mastroianni, M., Wizmer, E., & Amodeo, V. (2013). Teaching science through pictorial models during read-alouds. Journal of Science Teacher Education, 24(2), 367–389.

Oliver-Hoyo, M., & Sloan, C. (2014). The development of a visual-perceptual chemistry specific (VPCS) assessment tool. Journal of Research in Science Teaching, 51(8), 963–981.

Opfermann, M., Schmeck, A., & Fischer, H. E. (2017). Multiple representations in physics and science education—Why should we use them? In D. F. Treagust, R. Duit, & H. E. Fischer (Eds.), Multiple representations in physics education (pp. 1–22). Cham: Springer.
Pence, H. E. (2010). Smartphones, smart objects, and augmented reality. *The Reference Librarian, 52*(1–2), 136–145.

Qassem, L. M. M. S. A., Hawai, H. A., Shehhi, S. A., Zemerly, M. J., & Ng, J. W. P. (2016). AIR-EDUTECH: Augmented immersive reality (AIR) technology for high school Chemistry education. In *Proceedings of 2016 IEEE Global Engineering Education Conference (EDUCON)*, 10–13 April 2016, Abu Dhabi, UAE (pp. 842–847). Piscataway, NJ: IEEE. https://doi.org/10.1109/EDUCON.2016.7474650

Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review, 29*(4), 717–761.

Schnott, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and instruction, 48*(2), 141–156.

Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles, 53*(5), 433–441.

Tregast, D. F., Chittleborough, G., & Mamiula, T. (2003). The role of submicroscopic and symbolic representations in chemical explanations. *International Journal of Science Education, 25*(11), 1353–1368.

Tregast, D. F., & Tsui, C.-Y. (2013). Conclusion: Contributions of multiple representations to biological education. In D. F. Tregast, & C.-Y. Tsui (Eds.), *Multiple representations in biological education* (pp 349–367). Dordrecht: Springer.

Urahne, D., Nick, S., & Schanze, S. (2009). The effect of three-dimensional simulations on the understanding of chemical structures and their properties. *Research in Science Education, 39*(4), 495–513.

van Bruggen, J. M., Kirschner, P. A., & Jochems, W. (2002). External representation of argumentation in CSCL and the management of cognitive load. *Learning and Instruction, 12*(1), 121–138.

Wan, A. T., San, L. Y., & Omar, M. S. (2018). Augmented reality technology for year 10 chemistry class. *International Journal of Computer-Assisted Language Learning and Teaching, 8*(4), 45–64. https://doi.org/10.4018/IJCALLT.2018100104

Wu, H.-K., & Krajcik, J. S. (2006). Inscriptional practices in two inquiry-based classrooms. A case study of seventh graders’ use of data tables and graphs. *Journal of Research in Science Teaching, 43*(1), 63–95.

Wu, H.-K., & Puntambekar, S. (2012). Pedagogical affordances of multiple external representations in scientific processes. *Journal of Science Education and Technology, 21*(6), 754–767.

Wu, H. K., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education, 88*(3), 465–492.

Yang, E.-M., Andre, T., Greenbowe, T. J., & Tibell, L. (2003). Spatial ability and the impact of visualization/animation on learning electrochemistry. *International Journal of Science Education, 25*(3), 329–349.

Yu, S.-J., Sun, J. C.-Y., & Chen, O. T.-C. (2017). Effect of AR-based online wearable guides on university students’ situational interest and learning performance. *Universal Access in the Information Society, 8*(6), 1245.

Zheng, M., & Waller, M. P. (2017). ChemPreview: An augmented reality-based molecular interface. *Journal of Molecular Graphics and Modelling, 73*, 18–23. https://doi.org/10.1016/j.jmgm.2017.01.019

Zimmermann, H. T., Land, S. M., & Jung, Y. J. (2016). Using augmented reality to support children’s situational interest and science learning during context-sensitive informal mobile learning. In A. Peña-Ayala (Ed.), *Mobile, ubiquitous, and pervasive learning* (pp. 101–119). Cham: Springer.