Real Users, Real Results: Examining the Limitations of Learning Styles Within AEH

Elizabeth Brown¹, Tony Fisher², Tim Brailsford¹
¹School of Computer Science & IT | ²School of Education
University of Nottingham
Wollaton Road
Nottingham
NG8 1BB

{elizabeth.brown, tony.fisher, tim.brailsford}@nottingham.ac.uk

ABSTRACT
This paper examines the current state of AEH (adaptive educational hypermedia) research into explicit learning style modelling for user personalisation. It addresses the problem of non-naïve test subjects, who are often in user trials, thus contributing to experimental bias. Instead, the authors suggest using “real people”, i.e. users with a range of backgrounds and abilities, in order to gain a truer insight into evidence-based research.

We report on a study carried out with “real” users: around 80 children at a UK primary school. The study investigated sequential and global learning styles as a personalisation mechanism in an AEH system. The user trial involved matching and mismatching users and learning environments to see if learning improved. The AEH system used by the children was DEUS, a new e-learning platform that is conceptually similar to WHURLE, an AEH that also used learning styles as its user model.

No statistically significant differences were found between experimental groups, learning style preferences or learning environments. We discuss the significance of this, and then critically analyse the use of learning styles in relation to this study and also in the wider context.

Categories and Subject Descriptors

General Terms

Keywords
Adaptive educational hypermedia, learning styles, user modelling, DEUS, user trials, real users.

INTRODUCTION
1.1 Existing AEH systems
Adaptive educational hypermedia (AEH) frameworks have evolved over the last two decades in response to a growing need to provide personalised e-learning. A “one-size-fits-all” approach is no longer seen as desirable, since it is pedagogically limited and does not address the diverse range of learners’ preferences or needs.

The way in which personalisation is carried out is currently a thriving area of hypertext research across the world. Some adaptations might be controlled via the link level (as seen in the adaptive navigation support in Interbook, for example [7]), or to the content (utilised by systems such as AHA [13], MOT [12] and WHURLE-HM [30]). Other forms of personalisation might be carried out via structural adaptation [19, 25] or presentational adaptation [29].

Often, user modelling concentrates on the users’ prior knowledge. Some researchers regard this as too primitive in identifying a user’s needs, and thus different user models are being investigated. One of the most enduring notions of recent times is that of learning styles as a way of providing users with information tailored to their needs. A plethora of publications point to their efficacy in improving learning [3, 4, 8, 18, 24] but many of these studies cannot be considered to be scientifically rigorous. Several studies do not have satisfactory control groups (or at least a crossover design); most of them have insufficient sample sizes; nearly all do not have a long enough intervention period and none of them seem to make mention of the Hawthorne effect (akin to a placebo effect, whereby the behaviour of test subjects is altered because of their participation in the study and not necessarily caused by any specific intervention).

In addition, an extremely comprehensive report published by Coffield in 1994 [11] states that many learning style models are flawed, and do not possess adequate reliability or validity to measure what they are intended to.

Another problem often encountered by AEH researchers is that of their test subjects. It is likely that many researchers do not even realise that this is a problem – many will simply use whatever participants they have easy access to.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HT'07, September 10–12, 2007, Manchester, United Kingdom.
Copyright 2007 ACM 978-1-59593-820-6/07/0009...$5.00.
1.2 The problem with participants
Too many studies of AEH systems are tested on what are colloquially termed “pet students”, i.e. users who are often targeted as test subjects due to the ease of access to those students and the convenience of the experimental staging. These test subjects are usually taught students (or sometimes research students) to whom the authors of the study interact with in the course of their teaching commitments or research. Unfortunately, these students are often studying at university level and so cannot be called genuinely naïve users, since they have already been pre-selected on intellectual ability. In addition, they already have the capacity to learn effectively from a number of resources.

Instead, we propose that users for such studies should be based in the “real world”; they should be learners who do not know how to learn effectively already and there should be no pre-selection on general intellectual ability. The resultant data will therefore be more realistic and more widely applicable than studies that just use university students as their test subjects. It would also, theoretically, be of greater benefit to its “real world” target users since they would, by definition, not necessarily be able to learn effectively already. Working on the supposition that a personalised hypermedia system would improve learning, this could be an invaluable additional resource for teaching, both inside and out of the classroom.

1.3 AEH use in schools
One such group of “real” users are school children. They are a diverse group of learners with a wide range of intellectual (and other) abilities. Moreover, although some work has been done with school children and learning styles, hardly anything has been done with school children and AEH.

Bajraktarevic et al carried out 2 studies with GCSE pupils (aged 14-15 years old) [2, 3] that investigated learning styles within AEH. Using the ILASH system, summarising and questioning strategies were supported by link and text adaptation in the user interface. However, an evaluation of this system has not yet been published so it unknown whether this aided students’ learning or not. In their other study, Bajraktarevic and her colleagues used global and sequential learning styles as the adaptation mechanism in a web-based learning system. They found that students in a matched session achieved statistically significantly higher results than students in a mismatched session. Browsing times were not affected, however: students in both sessions spent approximately the same amount of time browsing, regardless of their learning styles being matched or mismatched to the learning environment.

Whilst these two studies are extremely interesting, and should be applauded for their efforts to involve “real users” in the user trials, subsequent research carried out with schools seems to have been neglected. In particular, the authors are not aware of any quality research that has been carried out with primary school children, who have even less experience of learning than GCSE pupils, and might well benefit from the support that can be provided by an AEH system.

1.4 Overview of case study
A detailed study was carried out at a local Nottinghamshire primary school, in consultation with the headteacher and staff. The aim of the study was to investigate matching and mismatching pupils’ learning styles with a web-based AEH system. Initial investigations were carried out to determine the range of pupils’ learning styles, and discussions were carried out with the class teachers to ascertain the topics that the pupils would study throughout their use of the system.

The AEH system was trialled with 5 sample users before being used in the main phase of the study with pupils from Years 5 & 6 at the target primary school. The school involved in the study is a primary and nursery school in Nottingham and has around 300 pupils on roll. It is a mixed school, taking pupils from 3-11 years old and is non-denominational. The majority of pupils are from white British backgrounds, with English as a first language. The proportion of pupils with learning difficulties and/or disabilities is close to the national average for the UK. Attainment on entry to the school is slightly below national expectations.

To summarise, this is a typical UK primary school with a wide variety of pupils showing a range of characteristics, preferences and abilities: an ideal place in which to find examples of “real users”.

1.5 Choice of learning style
In previous work, the authors implemented a visual-verbal learning styles filter for the user model in an existing AEH system [5, 6]. However, this suffered from several problems:

- The majority of people have a visual learning preference, to a greater or lesser degree\(^1\); hence adapting for verbal learners is not cost-effective.
- It can be difficult to determine the granularity of how strongly visual a picture or diagram is.
- The visual representation of some concepts can be subjective and susceptible to bias by the content author.
- Visual and verbal information equivalence can be a major problem across many knowledge domains.
- Paivio’s dual coding theory [23] states that people learn most effectively from a mixture of text and pictures.

Due to these reasons, and the fact that no statistically significant benefit could be found by matching users’ learning style preferences with their learning environment, it was felt that this user model was of limited value and should not be used for further research of this nature\(^2\).

Instead, it was decided to utilise global and sequential learning styles as the user model, in a similar manner to Bajraktarevic [3]. This is because it is a learning style that is common to many models, although sometimes referred to “wholist” or “holist” rather than “global”, and “analytic” or “serialist” in place of “sequential” [9, 26]. It also translates easily to hypertext and hypermedia structures and provides a logical mapping of

---

\(^1\) As evidenced in this study, and supported by [15].

\(^2\) It could be argued that these studies suffered from using university students as test subjects, with the associated problems therein. However, the other reasons mentioned here are sufficient in themselves to justify abandoning this model.
information to various domains/concepts and/or navigational support. There are also several studies that showed positive results when this model was used with students [3, 8, 19].

Global and sequential learning styles relate to how the structure of information is presented to a user. Global users will prefer to receive information firstly as an overview (the “big picture”) before being given the finer details; they will tend to learn best by seeing how topics interconnect and relate to one another. Sequential users will prefer to receive small chunks of linear information, with details being presented first, followed by an overview of the subject. Sequential users can get lost if material is not presented in an orderly and systematic manner; they cannot cope very well if the information jumps around from one topic to another.

One distinct advantage to implementing this user model is that content need not be authored for multiple representations, as previously done for studies into visual and verbal learning styles. What is important here is the structuring of information – how it is presented to the user – not the use of graphics or text per se. Hence content need be created only once, but structured differently to meet the different learning style preferences of users.

2. EXPERIMENTAL DESIGN

2.1 Introduction

The experimental design for this study was carefully planned in consultation with colleagues from the School of Psychology at the University of Nottingham, and with the members of staff at the school.

The user trial was staged so that each pupil involved in the study would be exposed to an average 2-hour intervention period over a period of 2-3 weeks. Pupils would be assessed beforehand for their learning style and existing domain knowledge (the pre-test) and also at the end (post-test). They would then be manually assigned to an experimental group before using the system.

The intervention itself took the form of structured learning using information contained in the AEH system, where users would log in and browse the system and then answer questions in a workbook. The questions in the workbook were meaningless unless used in combination with the AEH system, and meant that pupils were given short, task-oriented activities designed to maintain concentration. Merely asking pupils to browse a system and learn in an independent manner is not appropriate for children of this age; it would also have introduced an additional compounding factor and according level of variance into others using the other environment, in terms of being able to access the required information more quickly. The design of the study was to investigate if pupils’ learning styles helped them or hindered them, in terms of browsing for information and learning specific concepts – hence homogeneity in other factors (such as information distribution and the look and feel of the interface) was paramount.

On completing the workbook, pupils were asked to complete the post-test for both learning style (to see if it had changed during the course of the study) and domain knowledge.

2.2 Pilot study

2.2.1 Data collection

Most of the commonly-used learning style models that have been integrated into AEH systems have not been used with users under 18 years old, with the exception of the Dunn and Dunn model [14]; other VAK-based (Visual-Auditory-Kinaesthetic) theories based on perceptual modalities and Gardner’s Multiple Intelligences [17]. This meant that assessing the global and sequential learning styles of pupils in the study would require modification of an existing tool.

One of the ways in which global and sequential learning styles can be assessed is through the Felder-Solomon Index of Learning Style (ILS) questionnaire [16]. The ILS questionnaire has 44 questions, that assess a person’s learning style on the following axes:

- Visual/verbal
- Active/sequential
- Sensing/intuitive
- Sequential/global

Each axis equates to 11 items in the questionnaire, each consisting of a question (e.g. “When I think about what I did yesterday, I am most likely to get…”) followed by 2 responses, one of which must be chosen by the respondent (e.g. “a) a picture; b) words”).

A modified version of the ILS questionnaire was piloted with 50 children. The main reason for the pilot was to ensure that the target users would be able to answer the questions, and would be able to understand what they were asking. It was also a chance to investigate the existing learning styles of the current pupils at the school.

The pilot questionnaire consisted of 2 main modifications from the original ILS one: reduce the number of questions to assess only visual/verbal and global/sequential axes; and alter some of the phrasings used so that children could understand the questions more easily.

The questionnaire was administered twice, with a 10-day time period between them. This was to determine the stability (or otherwise) of the learning styles under examination. The reason
that both visual/verbal and sequential/global axes were measured, rather than just the latter, was to provide additional data about school children’s learning styles and to see how their distributions compared to those of other learners.

### 2.2.2 Data analysis

From the pilot study, it is clear that the participants showed a mostly visual learning style preference; this is in keeping with other studies [15]. This did not change much over the time period; only 22% of pupils changed their overall visual-verbal preference from one time period to the other, although there is some small variation of the strength of the preference within each mode. The graph in Figure 1 shows the distribution of visual and verbal learning styles amongst pilot study participants; the left-hand side of the chart shows pupils with a visual learning style whilst those pupils showing a verbal learning style are represented on the right-hand side. The numbers on the X-axis represent scores from the modified ILS questionnaire: those scoring 9-11 at either end of the scale have a strong preference for that style; 5-7 represents a moderate preference and 1-3 indicates little or no preference. The majority of participants are on the left-hand side of the graph, indicating a marked preference for material to be presented in a visual manner.

![Figure 1. The distribution of visual and verbal learning style preferences in 10-11 year old children](image)

The data from this pilot study give us several important findings about the learning styles of 10-11 year old school children:

- visual/verbal learning style preferences tend to be skewed to the visual axis
- visual/verbal learning style is relatively stable over a short time period
- sequential/global learning style preference is normally distributed
- sequential/global learning style seems to be more changeable than visual/verbal learning style over a short time period

The normal distribution of sequential/global learning style means that there is more merit in using this as a user model than the visual/verbal learning style, since we can usefully cater for pupils at both ends of the continuum. Although there seems to be a large number of pupils grouped in the middle, where there might be only little or no preference for one learning style or another, there are still a sizeable number each side of this – these are the learners to whom we would wish to match the mode of information presentation, since they might not learn effectively otherwise.

### 2.2.3 Summary of pilot study

From interaction with pupils and their teachers at the time of the pilot, and subsequent data analysis, it is clear that participants in this initial study understood the questions and were able to give clear answers.

Having been tested thoroughly with a reasonable number of target users, the sequential-global items from this questionnaire were now ready to be used in the main phase of the study.

### 2.3 The creation of DEUS

Previous work carried out by the authors into user modelling in AEH used a system called WHURLE (Web-based Hierarchical Universal Reactive Learning Environment). WHURLE is an
XML/XSLT system designed to deliver adaptive content over the web. The content is contained in a series of small units of information called “chunks” and a lesson plan determines which chunks are presented to each type of user in the user model. WHURLE has a pluggable user model, which can be switched interchangeably according to the study. Existing user models include a hybrid design, that assess a user’s prior knowledge within a specific domain [31]; and a learning styles model that implements visual/verbal learning styles [5]. Although it would have been fairly trivial to modify the existing learning styles model to include sequential/global preferences, it was felt that the system architecture was needlessly complex for the proposed user trial. In addition, page rendering was often slow in previous studies, and this was seen as highly undesirable when working with younger users.

A new A EH system was proposed, one that would be conceptually similar to WHURLE in terms of its lesson plan and content chunks, but much simpler and quicker to access.

The new system, DEUS (Digital Environment Utilising Styles) was coded in PHP and MySQL. User profiles are stored in the mySQL database and managed via PHP. The lesson plan and resulting navigation system was managed via third-party software: xhawk phpNav (http://xhawk.net/projects/other). Xhawk phpNav creates a customisable navigation system from a text file that is rendered in XHTML. In DEUS, 2 text files were used: one for a sequential lesson plan and another for a global lesson plan. The lesson plan contained details of what chunks should be presented to the user, and also the ordering of the chunks. Chunks were units of taught information consisting of a mixture of images, text, movies and Flash animations, written in XHTML. They were incorporated into the interface via PHP include statements.

The interface was designed to be as simple and intuitive as possible, with only 3 main parts to it: a horizontal frame at the top (containing the system logo and the user’s name); a left-hand navigation system; and a right-hand frame that contained the content (see Figure 3 below).

![Figure 3. Screenshot of DEUS, showing the navigation system on the left and example content on the right](image)

Users had individual logins to DEUS and when using the system for the first time, were prompted to answer pre-test questions addressing learning style preference and existing knowledge. There were 11 learning style questions and 30 multiple choice questions to test knowledge.

There was a week’s time delay between this first login and subsequent logins, to allow researchers time to manually assign pupils to experimental groups. This could have been done automatically by a PHP script, but it was felt that manual assignment to groups would help control for additional factors such as age and gender, providing better homogeneity between the experimental groups.

### 2.4 Modelling learning style in DEUS

The system architecture of DEUS was intentionally simple: a repository of content chunks, together with some rules that decide which chunks should be presented to users, and in what order.

The learning styles adaptation mechanism is essentially contained within the lesson plan, which is itself related to the navigation structure.

Each lesson plan consisted of several topics within the main unit being studied (“Life cycles”). Each topic consisted of a list of sub-topics, each one containing several linked pages. However, the global lesson plan had an additional first topic, which gave an overview of the other topics; this was not provided in the sequential lesson plan. Information overviews were given in the sequential lesson plan, but only at the end of each topic, to summarise the details previously presented to the user.

To a casual observer, the interface does not look very different when viewing either the sequential or global lesson plan. This was an important part of the study, since a perceived difference in the interface would introduce greater variance to the experimental design. By controlling this factor, any differences between experimental groups could not be affected by it.

It would be easy enough to integrate other learning style models into the DEUS system architecture, since all that is required is the modification of the lesson plan – and possibly some additional content creation, depending what the user model is and whether the existing content would be appropriate.

### 2.5 Choice of content

The unit chosen for the pupils to study was decided upon in conjunction with the teaching staff at the school. “Life cycles” was chosen since most pupils would not have studied the topic before and so it would be new to many of them. In addition, it was a topic that all pupils would study later in the year, and so by exposing them to the concepts in this study, it was hoped that the information would be retained and then reinforced more effectively.

This National Curriculum Key Stage 2 topic essentially teaches pupils about reproduction in flowering plants; reproduction in humans (at a very basic level); the concepts of extinction and the consequences of not reproducing.

The school governors gave permission for the study to go ahead; this was especially important due to the delicate nature of some of the subject matter. Parents/guardians were informed of their child’s participation in the user trial ahead of the study, and were
given the opportunity to withdraw this participation, although none did.

2.6 Additional details of the user trial
The workbooks used by the pupils were created in both global and sequential modes, to be used alongside the associated lesson plan. Users were given visual prompts on screen when they were required to answer questions in their workbook about the content that they were browsing. Pupils were offered prizes for the best workbook, thus maintaining a high level of motivation throughout the course of the study.

Throughout the study, the majority of the toolbars in the pupils’ web browsers were turned off, including the address bar. The reasons for this were twofold: firstly, to prevent the children from manipulating the address bar and ending up with page errors; and secondly to maximise the size of the interface on screen, since the computers at the school had only 14-inch CRT monitors. These measures were put in place following formative feedback of initial testing of the system by other target users prior to the study.

Only 12-13 pupils could participate in a particular session due to the limited number of working computers in the school’s ICT suite. Due to timetabling issues and other school activities, a maximum of 5 sessions could be conducted on any one day. On average, researchers carried out 3 sessions per working day.

3. DATA ANALYSIS
82 pupils from Years 5 & 6 (aged 9-11 years old) participated in the study over a 3-week period. Once assessed for learning style and existing knowledge in the pre-test, they were assigned into one of 4 groups: matched sequential; mismatched sequential; matched global and mismatched global. Group sizes were unequal in size, but with a minimum number of 15 pupils; this is the smallest sample size that should be used for any grouping within an experimental situation ([22], cited in [28]). Groups were created as homogeneously as possible, with an equal mix of ages and genders.

The experiment sought to address the issue of matching or mismatching; would there be any statistically significant benefits to matching users’ learning styles with their environment? Or, conversely, any disadvantage? These and the variety of other hypotheses under investigation are addressed in the next section.

Assessment was carried out by pre-test and post-test quizzes. Pupils’ workbooks were also marked and given back to them, but this feedback was qualitative more than quantitative, and did not form part of the formal measures used to assess pupils’ knowledge gain. Individual user tracking was carried out via web access logs and these were analysed to provide evidence to support or refute the various hypotheses.

3.1 Main hypotheses
The key focus of the study was to find out whether there was any benefit (or disadvantage) to matching or mismatching users’ learning style preferences to their environment. We also wanted to investigate if one learning style or learning environment was better than another in terms of the knowledge gained. The null and alternate hypotheses can thus be represented formally as a series of statements.

H₀ – there will be no statistically significant difference in knowledge gained between users from different experimental groups
H₁ – students who learn in a matched environment will learn significantly better than those who are in mismatched environment
H₂ – students who learn in a mismatched environment will learn significantly worse than those who learn in a matched environment
H₃ – one particular type of learning style provides more effective learning than another
H₄ – one particular type of learning environment provides more effective learning than another.

The measure used to assess knowledge gained was the post-test quiz score minus the pre-test quiz score. The data related to these hypotheses were all analysed using univariate Analysis of Variance (ANOVA); the results from these statistical tests are presented in the following section.

3.2 Experimental results
Table 1 below shows the average difference in knowledge gained between matched and mismatched users. Matched users were those who learning styles were matched with the lesson plan e.g. sequential users with a sequential lesson plan. Mismatched participants used a lesson plan that was not matched to their learning style, e.g. sequential users with a global lesson plan.

The two matched and two mismatched sub-groups were aggregated into 2 larger groups respectively.

<table>
<thead>
<tr>
<th>Group type</th>
<th>N</th>
<th>Avg. knowledge gained (points)</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matched</td>
<td>41</td>
<td>1.32</td>
<td>3.297</td>
</tr>
<tr>
<td>Mismatched</td>
<td>41</td>
<td>1.93</td>
<td>3.670</td>
</tr>
</tbody>
</table>

Statistical testing was carried out using SPSS, to determine if there was any significant difference in knowledge gained. However, the difference between the two groups was not significant ($F(1,80)=0.626, p=0.43$) and hence we can reject hypotheses 1 and 2.

When taking into account learning style and learning environment type as separate factors, there did also not seem to be any significant differences between groups (see Tables 2 and 3 below).

<table>
<thead>
<tr>
<th>Learning style</th>
<th>N</th>
<th>Avg. knowledge gained (points)</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential</td>
<td>52</td>
<td>1.4</td>
<td>3.195</td>
</tr>
<tr>
<td>Global</td>
<td>30</td>
<td>2.0</td>
<td>3.957</td>
</tr>
</tbody>
</table>
Since there was no statistically significant difference between pupils from different learning style groups in terms of knowledge gained ($F(1,80)=0.55$, $p=0.46$), there is clearly no benefit to possessing one learning style over another.

In addition, there did not seem to be any benefit to using a particular type of learning environment, since the differences between groups was also not significant ($F(1,80)=0.51$, $p=0.48$). Hence we can also reject hypotheses 3 and 4.

Another aspect of the study was to investigate any differences in knowledge gained between pupils with Special Education Needs (SEN) and those who did not have Special Educational Needs, in case this created complications in any of the sub-groups that might have had a disproportionately higher number of SEN pupils. This was formally represented in an additional hypothesis:

$H_5$ – SEN pupils had different levels of academic achievement from non-SEN pupils.

From Table 4, it can be seen that although there was some difference in knowledge gained between the two groups, this difference was not statistically significant ($F(1,80)=1.62$, $p=0.21$). Hence the proportion of SEN pupils to non-SEN pupils within different experimental groups would not have had any covariant effect on any of the previous data analyses.

### Table 4. Difference in knowledge gained between SEN and non-SEN pupils

<table>
<thead>
<tr>
<th>Type of pupil</th>
<th>N</th>
<th>Avg knowledge gained (points)</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEN</td>
<td>35</td>
<td>1.06</td>
<td>3.134</td>
</tr>
<tr>
<td>Non-SEN</td>
<td>47</td>
<td>2.04</td>
<td>3.695</td>
</tr>
</tbody>
</table>

#### 3.3 Summary of main findings

From the statistical analysis carried out of the main hypotheses, we can see that there is no evidence to support the theory that matching users’ learning styles to learning environment has any impact on making learning more effective. Nor does possessing a particular learning style (regardless of environment) make any difference – pupils with global learning styles seem to perform the same as pupils with sequential learning styles. When taking into account the learning environment, regardless of learning style, there is still no difference between groups. We therefore reject hypotheses 1-4 and instead revert to the null hypothesis, which is supported by these findings, namely that there is no statistically significant difference in knowledge gained between matched and mismatched users.

The authors also examined the post-test quiz score independently of pre-test score, in relation to all of the above hypotheses, but the difference between groups failed to reach significance for any of them.

### 3.4 Additional findings

In keeping with previous studies [3, 18], it was also decided to investigate different amounts of time taken to browse the system, since this could be used as a marker of how effective the system is to learn from, e.g. if someone can learn the same amount in a shorter time then this is better than someone who spends longer on it for the same academic gain.

We examined several hypotheses, and again analysed the data with univariate ANOVA tests:

- $H_6$ – there will be no statistically significant difference in amount of browsing time between users from different experimental groups
- $H_1$ – browsing times are different in matched versus mismatched users
- $H_2$ – browsing times are different for different learning styles
- $H_3$ – browsing times are different for different learning environments
- $H_4$ – browsing times are different for SEN pupils compared to non-SEN pupils.

Browsing time was recorded implicitly by the number of sessions required by users to complete the workbook task. The number of sessions varied between 2 and 6, but averaged at nearly 4 per user. This equates to approximately 2 hours of intervention time (30 minutes per session).

Based on the findings of the study so far, researchers did not expect to find any differences between any of the groups. This was certainly the case for the hypotheses 1 and 2: there was no statistical significance between matched or mismatched groups ($F(1,80)=0.61$, $p=0.44$), nor between users of different learning styles ($F(1,80)=1.097$, $p=0.298$). This supports the findings of Bajakturevic et al [3].

However, a statistically significant difference was found between users of different learning environments ($F(1,80)=4.522$, $p<0.05$). Table 5 shows details of the average number of sessions attended by users from each experimental group.

### Table 5. Difference in number of sessions in users from different learning environments

<table>
<thead>
<tr>
<th>Environmental type</th>
<th>N</th>
<th>Avg number of sessions</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential</td>
<td>39</td>
<td>4.08</td>
<td>1.085</td>
</tr>
<tr>
<td>Global</td>
<td>43</td>
<td>3.56</td>
<td>1.119</td>
</tr>
</tbody>
</table>

The data in Table 5 shows that pupils studying in the sequential learning environment took slightly longer to browse the system than those using the global learning environment. Since the amount of content contained in each environment is the same, merely ordered differently, this variation cannot be explained by suggesting there is more content. Nor is this an artefact of uneven information distribution, since this was controlled for in the design of the study.

The final additional hypothesis examined differences in browsing times between SEN and non-SEN pupils; however, no statistically significant differences were found ($F(1,80)=1.339$, $p=0.25$).
From these additional hypotheses, it can be seen that there is partial support for hypothesis 3, so the null hypothesis can be rejected. Hypotheses 1, 2 and 4 can also be rejected since they are not supported by the data.

3.5 Correlations

Lastly, we examined the data for any sensible correlations, using Pearson product-moment coefficients. There were no correlations between knowledge gain and browsing time ($r=0.05, n=82, p=0.65$) meaning that pupils did not learn more if they spent more time browsing the system.

However, there was a significant negative correlation between post-test score and browsing time ($r=-0.364, n=82, p<0.001$). This indicated that the longer the user spent browsing the system (and completing the workbook), the lower the post-test score. The authors postulate that this could be attributed to a motivation effect: that the pupils taking longer to finish the workbook were getting de-motivated and bored, and this had a negative effect on aspects of learning.

There was also a significant (positive) correlation between the number of pages visited and the number of browsing sessions ($r=0.555, n=82, p<0.001$); this is only to be expected since the more time spent browsing, the more pages will be visited. This may at least indicate that pupils were browsing the pages for information to answer the questions in the workbook rather than making up the answers or copying them from the participant next to them. This is useful information in itself – it shows that pupils were on-task for much of the time, and engaged by the system rather than being distracted by external events.

4. CONCLUSION

From these experimental trials, there does not seem to be any clear evidence to suggest that using global and sequential learning styles as a means of personalisation in AEH has any clear educational benefit to users. It is possible that by providing an introduction to the material, pupils managed to grasp the main concepts more quickly (as evidenced by the data shown in Table 5); however most teachers typically do this in the course of their everyday teaching and it is a common feature of much educational software. It is interesting to see that all pupils (on average) benefited from this overview, even when it was not regarded as appropriate to their preferred learning style. However, the authors are sceptical as to whether this is a true learning style phenomenon, or merely an artefact of the study. It is strongly recommended that further research be carried out before it can be definitively attributed to learning style, rather than any other factor.

A strong feature of the study was that the participants were what the authors term “real users”: a varied group of children who were not pre-selected for the study on any criteria. There was a mix of ages, abilities, backgrounds and gender. This was, in every sense of the phrase, “real world research” and it is fair to assume that the findings of this study could reasonably expect to be mirrored elsewhere. We would encourage further studies of comparable users to contribute further evidence to support or refute our findings, since they are mostly at odds with the positive results shown by other AEH studies.

This study, however well-intentioned and designed, was unfortunately not without its flaws. One of the main problems encountered was how to assess the learning styles of 9-11 year old children, since there are no freely-available tools to do this. The resulting questionnaire produced for this study, although based on an existing document, was not tested for internal validity and reliability and hence the researchers introduced some slight variance right from the start of the study. The modified questionnaire was created in consultation with an HCI specialist at the University of Nottingham, so that the new questions were as equivalent to the original ones as possible; however this still introduces a possible source of error into the study.

In addition, although our pilot study shows that sequential-global learning style can be fairly changeable, this dynamic nature of user preferences was not catered for by the system. The test-retest reliability score in our main user trial, comparing pre-test and post-test scores for learning style, was low: $r=0.105, n=82, p=0.361$. This correlation was achieved when pupils’ raw scores for their learning style were analysed from the start and end of the study (scores ranged from 11 to -11 for each pupil). However, even when pupil scores were recoded into simply ‘sequential’ or ‘global’, the resulting correlation was still low ($r=0.138, n=82, p=0.232$). This suggests that learning style has a relatively low stability and throws into greater doubt the usefulness of adapting the teaching/learning experience to such a dynamic user characteristic. It also could have had an impact on the outcomes of this study. However, it is worth noting that the post-test learning style assessment indicated that 62% of pupils in this study had the same learning style at the end of the study as at the start; it is not known if this changed in the period in-between though.

The study also did not control for the Hawthorne effect, although in reality this can be rather difficult to do. In order to control for this effectively, participants in the study cannot realise that they are part of a study where they might be being given special treatment. This is not always possible to do, and so the best way to deal with this is to use a control group, or have a crossover design (where all participants experience the different interventions, at different phases of the study). However, the Hawthorne effect is more of a problem when positive results are observed, since it is easy to attribute causality to the intervention rather than to the effect of being observed.

The authors are of the opinion that there are 2 overarching problems with studies such as these, although the first is related to the second.

The first problem is that of learning styles. Learning styles is a contentious issue with many psychologists and neuroscientists, who have questioned their scientific basis and the theories that the models are based upon [10, 20]. The work of Coffield et al [10, 11] suggested that many learning style models do not in fact measure what they are intended to, and have low internal reliability and validity. The Coffield reports still do not seem to have had the international recognition that they deserve (although perhaps this is understandable from some parts of the world, where the report could be damaging to certain vested interests). As a result, it seems that many AEH researchers are still naively using learning styles as user models for their studies, which are generally flawed in their experimental design in some way. Thus, when positive results are found, it is easy to think that learning styles have a significant impact on learning, whereas in fact, the real reasons for the positive results are almost impossible to determine.

The second problem is that which is common to educational research: the process of learning. Learning and the factors that affect it are extremely complex; there are many different
influences that together or separately affect how people learn. Issues such as overall IQ (itself influenced by several factors); motivation; socio-economic background; time; effort; health (or lack of it); other distractions related to personal and social activities or commitments – these all contribute to how, when and under what circumstances somebody learns. A phenomenon commonly seen in schools is the effect of the weather on pupil behaviour. Many experienced teachers correctly predict that their pupils will become more badly-behaved when there is a sudden change in the weather (e.g. to rain or wind), as evidenced by Badger and O’Hare [1]. The nature of learning is obviously very complex, with a large interplay of factors. Learning (and the study of it) can therefore be said to be a “wicked problem”, a phrase coined by Rittel and Webber [27] to describe problems that are incomplete and contradictory, that often have changing constraints and resources. Solutions to wicked problems are often difficult to realise because of complex interdependencies, and cannot be considered finite: they tend to be “good or bad” rather than “true or false”. Although initially used by Rittel and Webber in the context of social planning, it seems to be a fitting terminology to use with studies of how people learn. Additionally, because of the complex nature of learning, it can be difficult to predict the outcome of certain interventions; even when the same intervention is carried out with in an identical study, a different result may occur.

With these two overarching problems, what then can be done from a pragmatic perspective to ensure future quality-assured studies?

5. FUTURE CONSIDERATIONS

It is clear that there is a need for high-quality AEH research, with well-designed experiments based on theoretically sound models. Fortunately, this is not difficult to do provided that the researchers involved are aware of these requirements. It is not expected that all researchers will necessarily have the prerequisite knowledge to design the experiments, or know to what extent the intervention needs to happen (or for how long), since this information can be gleaned from colleagues or existing literature. Many areas of computer science, including AEH, seem to blur into that of social science since much work is done with people – the users of our systems. An extremely useful guide to this kind of research is provided by Robson [28], who discusses the different kinds of experimental designs, including those that cannot be carried out in a closed laboratory setting. He states that the “gold standard” of applied research is the randomised controlled trial (RCT), where there is a control group and random assignment of users to either the control group or experimental group. He also discusses aspects of experimental design, such as sample sizes; data collection; interviews and questionnaires; and analysis of data.

Another aspect of high-quality research is to realise the limitations or controversies surrounding any part of the study, including the theoretical model upon which it is based. The learning styles model used as the basis for this study has been tested for reliability and validity, and found to have relatively high scores for both [21, 32] – hence its use here can be justified³. However, learning styles researchers should be aware of the problems that might be related to their particular choice of learning styles model.

Researchers should also think carefully on whom they should be testing their system. Obviously, if a system is created for a specific cohort of the population then they should be the target participants of any user trials. However, if a researcher were looking to generalise their findings to a wider perspective, they would do well to factor out any confounding variables, and think carefully whether the participants might have been non-randomly selected. In other words, are these “real people” that are being used, or just those that are convenient?

Lastly the authors suggest that another aspect of statistical analysis is considered, namely that of effect size. Effect size is not often mentioned, but the authors suspect that its exclusion from much of the published literature is detrimental to those studies. This is because, even when genuine positive results are found, it is difficult to tell the level of importance of those results. By calculating effect size, a more realistic viewpoint can be taken, since different interventions will have a common unit of comparison. In this way, phenomena that have a large positive effect size can be recognised as more useful than that having a small positive effect.

6. ACKNOWLEDGMENTS

Many thanks to the staff and children of Glapton Primary School in Nottingham for their enthusiastic voluntary participation in the user trials, with particular mention to Alison Coates (Year 5/6 co-ordinator) and Sheila Elliott (Headteacher). Also thanks to James Goulding, members of the Web Technologies Lab and Sue Jones (Mixed Reality Lab) at the University of Nottingham for their ongoing advice and support.

This research is supported by a PhD scholarship from the School of Computer Science & IT at the University of Nottingham.

7. REFERENCES


³ Although the modified questionnaire is open to valid criticism.


