CHAPTER 4

Evaluating the Effects of Media Representation

Since "Media and Method" (M&M) is the central concern proposed by this research to incorporate "intelligent media" to realize specific instructional design as well as effective Web-based learning. Therefore we divide the evaluation of CooTutor into two parts. The first part aims to evaluate the influence of *media*, i.e., the design of interactive 3D visualization as the cognitive media proposed by this research, while the second is to evaluate the *method*, which focuses on the influence of adaptive mechanism. In this chapter, we present the design of the experiment and preliminary results of the first part.

4.1 Design of the experiment

More precisely, we intend to answer the question mentioned in section 2.4 "*can we enhance spatial ability by using interactive 3D visualization?*" We adopt a media representation comparison approach, which is to investigate the comparative effects of using two distinct representations, i.e. 2D-based and 3D-based, on tutoring learners' spatial ability.

Note that the term "representation" could have various meanings. They mainly include two types of semantic in this thesis, "knowledge representation" used in AI systems or "media representation" used in discussing issues of educational media. Note that in this chapter, the term "representation" refers to "media representation". That is, the modality of symbol systems employed in instruction and learning, e.g., texts, pictures, 3D animations etc.

We use a pre-test/post-test comparison-group experimental design here, which is previously adopted by [16]. We designed two versions of learning materials with different representations but the same contents for the experiment. The difference of representations is thus the *experimental treatments* we employed. The overview of experimental design is shown schematically as follow.

Group 1:	S_1	R	T_1	S_2
Group 2:	S_1	R	T_2	S_2

R = random assignment of students
T_1, T_2 = experimental treatments, self-paced learning with different materials
S_1, S_2 = pre-test and post-test, both using the spatial ability test

The experiment was conducted in January 2004, at a computer classroom of National Taiwan Normal University. The duration of experiment was approximately 90 minutes[4]. Each participant has to complete a pre-test, the self-paced learning session within the system, and a post-test. A total of 23 undergraduate-level students, who are enrolled in a science education practice course at National Taiwan Normal University, involved in this experiment. Most of the participants are undergraduate students majoring in Earth Sciences. Six students among the participants are graduate students specialized in Science Education. We transplanted the *Purdue Visualization of Rotation Test* (PVRT) proposed by Bodner et al. [9] as a Web-based version. This Web-based PVRT was integrated into our system to assess all participants' spatial ability on both the pre-test and post-test.

The procedure of the experiment is expressed as an algorithm shown in Figure 4.1. At line 1 and line 2, when the student logged in the system all participants were randomly assigned to either the 2D group (*comparison group*, using 2D-based learning materials) or the 3D group (*experimental group*, using interactive 3D-based learning materials) by the system. At line 3 of the flow, the participants of both groups were tested with the identical Web-based PVRT immediately, as the pre-test step. Subsequently, at line 4, the experimental group was directed

[4] Since the learning progress inside most Web-based learning environments is controlled by the learner in nature, that is so called self-paced learning, it is reasonable to let learners decide the time to finish learning session and the time to enter the post-test. Therefore, the duration of experiment would vary depending on each learner's decision.

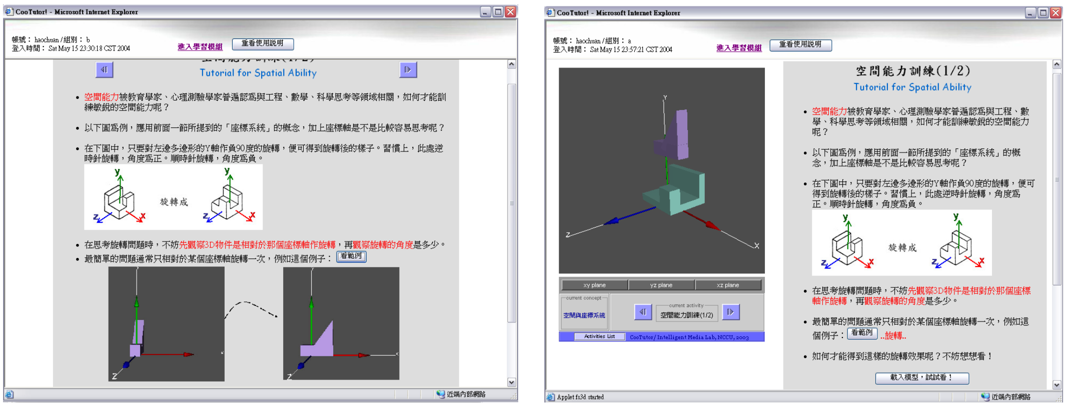
```
1  Participant P logged in
2  Pre-test
3  Assigning P to 2D or 3D group randomly
4  Learning session start
5      if P ask to finish the learning session
6          Ask P to make confirmation
7              if P confirms
8                  finish the session, enter the Post-test
9              else
10                 nothing to be done
11 Post-test
```

Figure 4. 1: The procedure of the experiment.

to the 3D-based version of learning environment whereas the comparison group was led to the 2D-based version. The learning environments for both groups were designed for self-paced learning. That is, participants could freely navigate learning materials the system offered, even repeatedly. Except for the difference of media representations, variables, such as contents, color style, and the number of pages is controlled to be identical for the two learning environments. Line 6 to 10 of the flow show the criteria for finishing the learning session. Participants of both groups could determine when to finish the learning session themselves. Once the participant intended to finish the session, this participant was prompted that the post-test will be conducted after the session, and asked to make confirmation. If she/he confirmed, the learning session will finish, and the post-test will be presented to the participant. Or if she/he regrets, then no decision would be done and the participant can freely navigate as usual. The design aims to enforce the participant to assess themselves meta-cognitively. The participant would have a chance to review her/his own learning status, and determine whether herself/himself is well-prepared. Finally, every participant was tested by the post-test.

4.2 Learning materials with different representations

In our experimental design, we intend to evaluate the influence of different modalities of representations on enhancing spatial ability. Two versions of learning materials were authored as mentioned previously. These two versions i.e. the 2D-based and 3D-based ones, both comprise seven HTML pages with identical contents. In the 3D-based version, the client-side modules of CooTutor, i.e., *tutor console* and *3D blackboard*, are employed. Participants of the



(a) (b)

Figure 4. 2: The screenshots of learning materials in different representations, (a) 2D-based version (b) Interactive 3D-based version.

3D group are allowed to manipulate 3D objects presented in the 3D blackboard. Functionalities of interactive 3D media are realized in the 3D-based version. On the other hand, for the 2D-based version, variables like color style and textual description was designed to be as identical to the 3D-based one as possible.

It is worth noting that the architecture of interactive 3D media allows the design of anchored instruction. That is, the student could press a button labeling "Press to show the demo" embedded in the Web page, and see the 3D animation demo presented in the 3D blackboard. We attempted to mimic this functionality, and embedded the same buttons in the 2D-based version. But in the 2D-based one, there was no 3D object and animation. After the student press buttons, previously hidden pictorial illustrations will become visible immediately. The screenshots of these two versions are shown in Figure 4.2, in which (a) is the 2D-based version, and (b) is the 3D-based version. Note that since we intend to make the comparison between media representations and to let students self-paced their learning in the experiment, there was no adaptivity in these two versions of environments.

4.3 Measuring instrument

To assess the students' spatial ability in our study, as we have mentioned, we use the Web-based PVRT transplanted from the version proposed by Bodner et al. in [9].

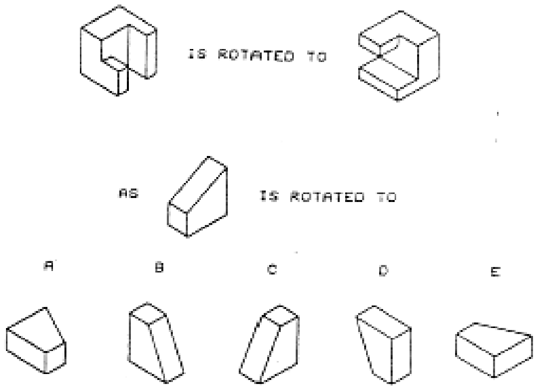


Figure 4. 3: A sample item of the 20-items version of PVRT, adopted from [9].

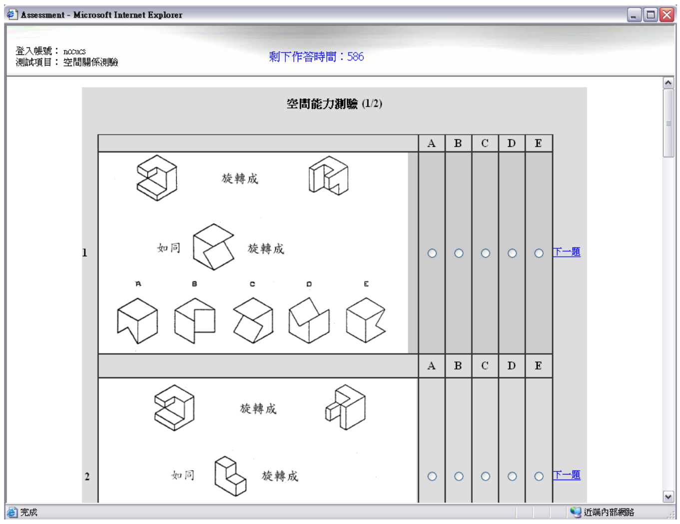


Figure 4. 4: The screenshot of Web-based PVRT.

The original version in [9] was deigned as a paper-pencil test, which comprises 20 items. In order to restrict the participants to use analytical answering techniques upon these test items, Bodner et al. also enforce a time limit of 10 minutes for the test. A sample item adopted from the paper-based version is shown in Figure 4.3. The estimated reliability coefficient of 0.78 to 0.80 by using Kuder-Richardson 20 formula (KR-20) was reported by [9].

The Web-based PVRT developed and used in this study consists of all 20 items of the paper-based version. For the convenience of answering, these 20 items were laid out as two HTML pages where each page shows 10 items. The sequence of items conformed to the paper-based version. Also, the time limit of 10 minutes (i.e. 600 seconds) was enforced strictly. A timer was shown to participants for reminding the rest of time. Figure 4.4 shows the screenshot of Web-based PVRT.

In order to prevent participants from answering the post-test by their impressions in pre-test, we slightly change the order of choices for each item of the post-test. The Web-based PVRT was highly internal consistent and reliable on both pre- and post-tests. For the pre-test, the KR-20 coefficient was 0.80; and for the post-test, the coefficient of 0.88 was derived.

4.4 Data analysis

In order to examine and compare the effects of these two groups, we aimed to conduct an analysis of covariance (ANCOVA) on the post-test scores with the pre-test scores as the covariates. Thus, the independent variable of the analysis was the groups; the dependent variable was the post-test scores; and the covariate was the pre-test scores.

Before performing the analysis, several underlying assumptions of ANCOVA were first examined. They are the assumptions of *normality*, *homogeneity of variance*, and *homogeneity of within-group regression* [35][40]. The Kolmogorov-Smirnov tests revealed that both pre- and post-test scores for both two groups conforming to normality. The Levene's test of equality also assured that there was homogeneity of variances between two groups ($p>0.05$). However the result of homogeneity of within-group regression was not strictly conformed ($p=0.032$, smaller than 0.05). To decide the alpha level of statistical significance of this study, two factors are considered. First, because the experiment is a pilot study based upon a rather small sample size, it is viable to adjust the alpha level up. Second, due to the failure of passing the test of homogeneity of within-group regression, it would be necessary to tune the alpha level down to control Type I error properly as described Huitema in [39]. By balancing all these effects, the alpha level is set at $\alpha=0.02$ conservatively. The ANCOVA analysis was undertaken by using SPSS 8.0 (Statistical Package for Social Sciences).

Table 4. 1: Descriptive statistics of participants' pre- and post-test scores, number of participants: 2D group (n=10), 3D group (n=13)

Pre-test scores		Post-test scores	
2D Group Mean (SD)	3D Group Mean (SD)	2D Group Mean (SD)	3D Group Mean (SD)
13.80 (3.94)	15.08 (3.73)	12.80 (4.59)	15.46 (4.82)
Overall, n=23 14.52 (3.79)		Overall, n=23 14.30 (4.81)	

Table 4. 2: The results of ANCOVA by using the post-test scores as dependant variable with pre-test scores as covariate

Adjusted post-test scores		ANCOVA $F(1, 20)$
2D Group Mean (SEM)	3D Group Mean (SEM)	
13.24 (1.35)	15.12 (1.18)	1.10 ($p=0.307$, $\eta^2=0.052$)

4.5 Experimental results

Table 4.1 shows the descriptive statistics of participants' pre- and post-test scores assessed by the Web-based PVRT. Note that the score of the PVRT test represents the number of items the participant answered correctly. Since there were 20 items in the PVRT totally, the perfect score was 20. If the participant skipped any item of the test, then that item was treated as the participant answered incorrectly. An independent 2-tailed t-test was also conducted on pre-test scores to detect if there was statistical difference between the two groups. The result ($p>0.05$) implies that both groups were approximately at the same level of spatial ability at the pre-test.

As mentioned previously, the duration that each participant stayed in the learning session was also recorded. The data were, 2D group: Mean=4.13 minutes, SD=2.87 and 3D group: Mean=8.33 minutes, SD=4.03. An independent 2-tailed t-test was performed, and it shows

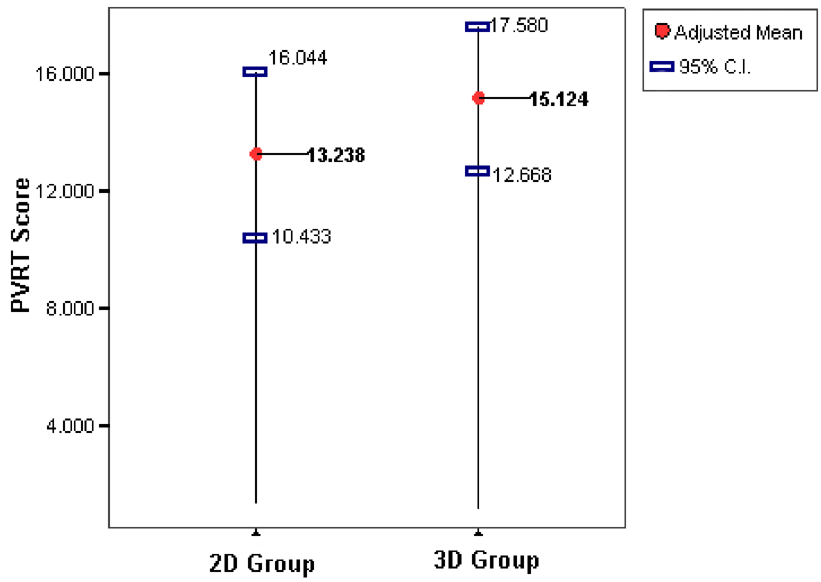


Figure 4. 5: The adjusted mean of post-test scores with 95% C.I.

that there existed statistical significance (p=0.008, smaller than 0.05). It could be inferred that participants of different groups were likely to perform distinct browsing behaviors.

The ANCOVA method allows us to eliminate the difference of pre-test scores between groups, and then the relative post-test scores could be derived. The adjusted post-test scores could reveal the real effects that the experimental treatment produced. Table 4.2 shows the result of adjusted post-test scores. The adjusted means with 95% confidence intervals of the result are also illustrated in Figure 4.5.

The adjusted post-test score of the 3D group is higher than the 2D group (2D: Mean=13.24, SEM=1.35, SD=4.27; 3D: Mean=15.12, SEM=1.18, SD=4.25). The result was *not* statistically significant, F (1, 20) =1.1, p=0.307 > α. Even so, it is worth noting that the overall mean of pre-test scores as shown in Table 4.1 was 14.52. While the 2D group's adjusted mean of post-test score is less than the overall mean of pre-test and the 3D group reveals an increase comparing to that value. The inclination between pre- and post-test looks quite different between 2D and 3D groups. In the next section, we will discuss this point and its implication further.

On the other hand, the *effect size* (η^2 =0.052, or f=0.234) according to the criteria of effect size proposed by Cohen (1988, p286) [18], is very close to a *medium effect size*. As McLean et al.(1998, p17) described , “the effect size gives an estimate of the noteworthiness of the results. [50]” The result derived via the analysis could be inferred that the practical significance is worth to note evidently.

4.6 Discussion and implications

In this section, the first issue to be discussed is the interpretation of the result of ANCOVA analysis. Researchers have noted the insufficiency of using only the result of statistical significance testing in statistical inference [18][23][50]. This is mainly because the computation of statistical significance is related to the sample size involved in the analysis. For a small size of samples, like this experiment, achieving statistical significance is inherently more difficult than for a large one. Inversely, it is quite possible to observe a statistical significance with a large sample size, even if there was little practical effect actually. Researchers even recognized that “an SST (statistical significance testing) is largely a test of whether or not the sample is large.”(Daniel 1998, p26) [23].

Statistical significance and practical significance (i.e., the coefficient of effect size) offer different aspects of information of the experiment. In this case, with a rather small sample size, the result of statistical insignificance is likely to happen in nature and thus less informative. And the result of medium effect size of practical significance is quite worth to note.

As mentioned previously, the experiment aimed to conform to the actual scenario of Web-based learning, and thus adopt an alternative design on offering students equivalent chance of learning. That is, it lets students to determine the time they want to leave the learning session. Since the factors underplayed students’ determination were not very clear, the interpretation of the statistical significance on learning duration should be undertaken cautiously. The difference may due to the functionality of interactive learning possessed by the 3D design. But it was not clear and beyond the scope of this study whether staying long is good or bad. On the one hand, this might be treated as an implication that students “like” 3D representations and are willing to spend their time to learn with the 3D. However, it is also possible that students felt perplexed by 3D navigation and thus spent more time on the task of manipulation. Roughly speaking, interactive 3D visualization will influence learners’ browsing behavior.

More experiments will be needed to address its implication

From the descriptive statistics shown in Table 4.1, it is interesting to note that there were inverse inclinations between the two groups. One may initially recognize that the post-test scores naturally become higher than the pre-test scores due to the effect of impression or getting familiar with the test. That is, so called practice effect in psychological experimentation [73]. However, this is not the fact observed in this study. For the 2D group, the mean of post-test scores is even worse than the mean of pre-test, whereas for the 3D group, the result shows a positive increase on post-test. The scenario happened to the 2D group was similar to the situation reported by Branoff et al. [8]. They note that participants involved in such pre-test/post-test experimental design may not give their best efforts on the post-test. Specifically, when the process of pre-test/treatments/post-test is contiguous within a short period of time (e.g., couple of hours), participants were easily to feel bored on the post-test. From the point of analysis, this might also reflect participants' attitudes. Since the treatment of the two groups was both conducted in the same classroom at the same period of time, the positive increase of post-test scores revealed by the 3D group is thus valuable to be noted.

The results from the ANCOVA analysis and the descriptive statistics provide a direction for research. The effect of using interactive 3D visualization in the Web-based learning context for spatial ability enhancement deems expectable. However, it is still unclear about the factors underlying the fuzzy enhancement. Besides the actual spatial ability possessed by students, it seems that students' attitudes toward learning activities also play a role on spatial ability test. The results suggest that it is necessary to conduct further replication studies on related issues.