Name: **M. Y. Carissa Young**

Degree: **MA in Information and Communications Technology in Education**
2000-2001

Module: **Research and Development in Information and Communications Technology in Education**
ICT.02C2

Coordinators: **Dr Andrew Brown & Dr Paul Dowling**

Title of Work: **A Critical Review of "To see a painting versus to walk in a painting: An experiment on sense-making through virtual reality" by A. Antonietti and M. Cantoia. C*omputers and Education*, *34* (2000), 213-223.**

Word Length: **3587 words**

1. Introduction

Reading research can be perceived as a one-way communication process between authors and readers. Since readers do not have the opportunity to ask authors for clarification while reading, communication breakdown might occur if a study has not been systematically constructed and reported. In this essay, I view an educational study (Antonietti & Cantoia, 2000) from the perspective of a reader using a mode of interrogation (Brown & Dowling, 1998). This essay addresses the extent to which the conclusions drawn by the authors can be justified on the basis of the experiments performed by focussing on three interrelated issues: the formulation of the problem, the design of the research, and the presentation of the results.

2. The Formulation of the Problem

The problem of Antonietti and Cantoia's (2000) study is explicitly presented as a hypothesis and later as a question. However, the structures of the hypothesis and the research question are not clear. Also, the concept variables presented in the problem are imprecise, which have made the research difficult to read.

2.1 The Structure of the Hypothesis

The study aims to test the hypothesis that "different cognitive activities should be elicited by a VR experience as compared to an instructional experience based on usual tools and do not produce an immersive contact between the

student and the content to be acquired" (p. 214). It is noted that the above hypothesis does not seem to have a clear syntactic structure. Neither does it conform to the format of hypotheses in the statistics convention (Robson, 1994). Had the authors wished to test the hypothesis by employing the statistical procedures reported later in the paper, they should have constructed a null hypothesis such as "there is no significant difference in 'cognitive activities' between exploring a virtual reality representation and a traditional representation".

2.2 The Readability of the Research Question

The research question that the study would address is "does the possibility to have a VR immersion experience, affect the act of making sense, by orienting thinking processes toward directions which differ from those hinted by traditional experiences?" (Antonietti & Cantoia, 2000; pp. 214-215). This research question does not appear to be legible, partly because of the ambiguous terms such as "affect" and "hinted", and partly because of the complex structure, which has obscured meaning. Again, the authors could have formulated a clearer research question like "Are there any differences in 'sense-making' between exploring a virtual reality representation and a traditional representation?".

The above analysis shows that both the hypothesis and the research question of the study are imprecise. Such less well-formed hypothesis and research question could not be accurately operationalised and eventually reconstructed

as conclusions. If these defects are purely linguistic, the authors could have dealt with them by seeking professional proofreading help. However, as discussed in subsection 2.3, there is a methodological weakness in the problem as well.

2.3 The Concept Variables

In spite of its unclear structure, the two concept variables in this problem could be interpreted as “exploration of sensory representation” and “sense-making processes” respectively. The former is a nominal scale with two values, namely “virtual reality” (VR) and “traditional”. “Virtual reality” is the essential concept of this study; however, as I will point out in the next section, the “virtual reality representation” tested when the research is operationalised differs from the concept presented in the problem. The other concept “sense-making processes” is defined as it is introduced. Yet the definition of “sense-making” as “a kind of situations, namely, cases in which the learner has to discover or construct a meaning, and attribute it to what is not front of him/ her” (p. 214) is by no means clear. “Making sense” cannot be equated to “situations” or “cases”. The term does not belong to the cognitive psychological lexicon either. It could be articulated as “construction of meaning”, which should be “processes”, not “cases” or “situations”. As discussed in subsection 3.4, the concept “sense-making processes” becomes somewhat different when it is stated as an indicator variable in the research design.

One possible reason for the ambiguity of the concept variables might be the unclear description of the problematic of the research, which is supposed to describe the specific area the authors are working within. However, the problematic of this research is constructed with reference to the characteristics of virtual reality and the advantages of this graphic representation only. The paper does not make adequate reference to the theory of learners' cognitive processes in virtual reality environments. Neither does it provide detailed information on the theory and practice of learning via exploring virtual reality objects. A review of the related studies on the relationship of virtual reality, cognitive processes and learning, which could provide readers with some rationale for learning in virtual reality environments, is not presented. In fact, the omission of a review of relative work on learning in virtual reality environments is surprising given its prominence in recent research. If the authors had made more reference to the literature on empirical research on computer-generated, interactive graphic representations and their potential use in learning and teaching, the concept variables could have been clearly defined. Yet the authors have failed to establish a clear problematic within which the research is situated. Since the problematic is not properly constructed, it is not surprising that the problem has not been successfully formulated. In this regard, it would seem that the defect of the problem could be not only a linguistic one but also a methodological one.

3. The Design of the Research

An analysis of the research design of this study has revealed more methodological weaknesses. This section focuses on four aspects: the sampling strategy, the concept-indicator coherence, the validity of the measures as well as the appropriateness of the experimental approach.

3.1 The Sampling Strategy

Regarding sampling strategy, the authors have correctly excluded those participants who had seen the research material in order to eliminate the chance that some participants had more/ fewer responses due to memory effects. However, they might not have properly controlled for other variables. As can be seen, the authors are trying to deploy a stratified sampling strategy by allocating equal number of participants with certain features to the two (VR and Reflection) conditions. In this research design, the variables balanced for are “gender”, “year of study” and “faculty”. The authors do not give any empirical evidence to justify their choice of these three variables. I would question, for example, why the authors control for “year of study”. Does a student’s year of study affect his/her “sense-making”? I wonder whether the authors have any empirical evidence to justify their choice.

Certainly the authors have the freedom to choose which variables to balance for, but they should select those which are the most important based on the nature of the research topic. In this study, the research topic involves the

comparison of two different spatial representations of objects, namely three-dimensional and two-dimensional pictures. An important, perhaps the most important variable which should be controlled for is *spatial ability*, which might "reflect the speed and ease with which basic cognitive processes are performed" (Anderson, 1995). Failing to do so, the authors might face this problem: some participants in the VR group might have higher spatial abilities and therefore processed the three-dimensional visual information faster and easier than others, hence the differences between the two groups. If my suspicion of the sampling strategy is proved to be right and the uncontrolled variable spatial ability does associate with the other variables, then the statistical tests reported later in the paper would be more difficult to interpret correctly.

3.2 The Concept-Indicator Coherence

The research material of this experiment is a painting of Saint Jerome by Antonello da Messina, presented as a computer-generated, three-dimensional, dynamic visual sensory representation for the VR group; and as a replicated, two-dimensional, static visual sensory representation for the Reflection group. It seems that the authors have attempted to have some control over the visual information that the VR participants were going to process by presenting to them a standardised "guided tour". However, I would argue that this computer-generated, three-dimensional standardised "guided tour" of a painting might not be a valid indicator of the concept "virtual reality representation" since it does

not have the features of three-dimensional aural representation and user's ability to manipulate information.

To support the above argument, I would like to refer to one of the papers cited by the authors (Ferrington & Loge, 1992), which describes the features of virtual reality. According to Ferrington and Loge, a possible reason why virtual reality may be of interest to educators and learners is the unique visual aural sensory experience it offers. For example, when moving closer to a VR representation of a bird on a tree, learners can have a closer look of the tree and hear a louder sound of the bird. In this respect, the "virtual reality" experiences tested in Antonietti and Cantoia's experiment, which only involve the visual sensory element, might not be the same as the "virtual reality" referred to by Ferrington and Loge.

The authors might well justify their claims by arguing that they were comparing two visual spatial representations of pictures, and therefore the aural input becomes irrelevant to the study. Nevertheless, I would challenge the authors: "why the VR participants were given a 'guided tour' and not allowed to manipulate the visual information spatially and freely?" Ironically, the authors seem to be aware of this feature, as they mention in the paper that virtual reality enables one to "change points of view in a flexible manner" (Antonietti & Cantoia, 2000; p. 214). However, in practice, they have never tested this interactive, or "flexible" nature of virtual reality. Again, it further reveals the weak concept-indicator agreement of the research.

3.3 The Validity of the Measure

I also have reservations about the use of a piece of fine art as a visual representation of three-dimensional and two-dimensional objects. My concern is that the different ways in which people perceive art representations might not be purely cognitive, but culture-based or even idiosyncratic. As stated by the authors, art representations are “polysemic in their own nature. Artistic products are open to multiple interpretations and different perspectives are available to analyse them” (p. 215). In this regard, I would question the extent to which “exploration of sensory representation of a painting” is a valid indicator of the concept “exploration of sensory representation”.

This weakness is related to another methodological concern: if the cognitive activities involved in processing visual art representations were culture-based, to what extent could the results obtained from the participants with certain cultural background be generalised to the population? An answer to this question is even more difficult to seek, since, in fact, the authors have never described in detail what the sample comprises and what population the participants of the study might represent. They have only reported in the paper that the participants of the study were forty student-volunteers from pedagogics and psychology departments at a certain academic institution. The inadequate information about the research participants could have weakened the claims that they later make when conclusions are drawn.

3.4 The Appropriateness of the Experimental Approach

Another weakness of this research could be the experimental design and the appropriateness of this experimental approach to the measure of the concept "cognitive activities" stated in the hypothesis, or "making sense" or "thinking processes" in the research question. I would question whether the experimental treatment, an experiment comprises four tasks, has properly measured any "sense-making processes". The tasks, which require the participants to write down the title of the painting, the meaning of the painting, the questions about the painting, and their comments on the painting, seem to measure certain cognitive outcomes, or more appropriately, the *products* of thinking. The experimental treatment thus carried out does not appear to examine the *processes* of thinking. Given this inappropriate experimental approach, the authors have merely measured *what* the participants know about the painting rather than *how* they construct meaning of the painting. They might not have adequate measures of the concept "sense-making processes" in this research. Since what the authors have measured appear to deviate from what they set out to measure, I wonder if such a research design might produce any results which allow the authors to successfully address the research problem.

I would like to propose an alternative experimental approach, which might have helped the authors measure the concept more accurately. As emphasised above, the authors' research interest is "thinking processes", not "thinking products". The former cannot be observed or measured directly because of its intrinsic nature. In order for researchers to infer to such processes, participants

should either verbally report on what they are thinking while performing a task or give verbal or written reports on what they have been thinking after task (Ericsson & Simon, 1984). Notwithstanding the ambiguous concept of “sense-making”, a more appropriate experimental design for the inference of “sense-making processes” could have been something that asks participants to give written reports on what they had been thinking while “making sense” of the VR/ static visual representation of the painting. An alternative design could have been to instruct the participants to report either introspectively or retrospectively on how they “make sense” of the painting.

4. The Presentation of the Results

The data collected from the study are classified, quantified and then statistically tested. However, the output of these procedures cannot accurately recognise the concepts of the research problem. This section critically reviews the construction of the classification scheme and the use of chi-square techniques in the study. Alternative approaches that the authors might have taken to present the research results are suggested.

4.1 The Construction of the Classification Scheme

The participants’ written responses to the experiment are presented as categories, which are classified “according to multiple criteria” (Antonietti & Cantoia, 2000; p. 216). The term “multiple criteria” is vague and a possible interpretation could be a certain classification scheme. According to the

Analysis of the Protocols section of the paper, such classification scheme might comprise four main categories, each of which has several sub-categories. Whilst the authors have used one example to describe each of the sub-categories, they have never explained how these categories and sub-categories are constructed. More importantly, the authors have never mentioned whether the classification system is established from scratch or it is derived from any theoretical or empirical work.

In fact, a more plausible presentation of the classification system could have been in the form of a table or diagram, in which each category is clearly labelled and defined. More appropriately, the authors would have given ample examples to illustrate the categories and sub-categories.

4.2 The Use of Chi-Square Techniques

The coded data are statistically tested using a series of chi-square tests and the inferential results are presented in the form of tables and text. While the authors claim that the statistical tests have yielded significant differences between the VR and the Reflection groups in some aspects of “sense-making”, they do not seem to know that the chi-square tests are used when the assumptions of these statistical procedures have been violated.

It would appear that the authors are not aware of the requirement that the chi-square techniques are appropriate for use with data in the form of frequencies only (Robson, 1994). In this research, the results are presented in the

percentage distribution of categorised responses by experimental condition. While the authors might have converted the data to counts before running chi-square tests, they should have reported the results more accurately by presenting the data in frequencies, which is appropriate for chi-square tests to be performed.

In fact, instead of presenting the statistical results in text, which has made the paper less intelligible, the authors could have used a table to illustrate the frequency distribution of category by experimental condition. If the chi-square techniques had been used appropriately, the authors could also have shown the tabulated chi-square values in the tables.

Even if the authors have converted the percentages to frequencies before using chi-square, an analysis of the raw data presented in this paper has revealed a second, and perhaps a more serious technical error: there is evidence that the chi-square tests are used when the assumption that the expected frequencies of all the cells should not be less that five (Robson, 1994) has been violated. To illustrate this point, the VR and Reflection participants' responses to the content of the title of the painting are converted to counts and shown in Table 1 (following Brown & Dowling, 1998; pp. 123-124).

Table 1: A Re-presentation of the Distribution of a Sense-making Sub-category of the VR and Reflection Participants in Antonietti and Cantoia's (2000) Study

Experimental Condition	Sense-making Sub-category: *Content of Title*	Observed Frequency	Expected Frequency
VR	Character	4.0	7.5
	Perceptual-spatial	4.0	3.0
	Abstract	8.0	7.5
	Meta-perspective	4.0	2.0
Reflection	Character	11.0	7.5
	Perceptual-spatial	2.0	3.0
	Abstract	7.0	7.5
	Meta-perspective	0.0	2.0

As can been seen, since the expected frequencies in four of the eight cells fall below five, chi-square test should not be used. Therefore, it is evident that the statistical procedures have not been appropriately performed in this research.

The above analysis has exposed the inappropriateness of the use of chi-square techniques in this study. I would question why the authors could employ a series of chi-square tests and how those "significant" chi-square values reported in the paper were tabulated when these three assumptions of using chi-square might have been violated. I would also question whether the statistical procedures employed in this study have adequately tested the experimental hypothesis that "different cognitive activities should be elicited by a VR experience as compared to an instructional experience based on usual

tools and do not produce an immersive contact between the student and the content to be acquired" (Antonietti & Cantoia, 2000; p. 214), which itself has been imprecisely constructed.

Regarding the use of statistics in this study, I have two suggestions. Had the authors insisted to use chi-square, they could have increased the sample size and/ or reduced the number of sub-categories so that the frequency distribution of participants in the sub-categories would have been higher. In this way, the expected frequencies could have been high enough in order for the test to be run. Yet they should have solved the problem of independent observation first. Alternatively, the authors could have converted the percentages to counts, and then performed the Fisher's exact test, which is appropriate for small-sample data in the form of frequencies (Robson, 1994). However, to use this test appropriately, the authors should have had confidence that the row and column totals had been fixed -- that there should be equal number of participants in each condition and that the count for each sub-category would be equal -- before running the test. Based on the nature of the research, it seems that this requirement is less likely to be fulfilled.

5. Conclusions

In this essay, I have dealt with three methodological concerns regarding Antonietti and Cantoia (2000). I have demonstrated that the problem of the research has not been constructed within a clearly identified theoretical

framework. Since the concept variables “sense-making processes” and “virtual reality” have not been accurately defined and the measures of them are inadequate and invalid, the concept-indicator coherence is weak. Under these circumstances, the data collected could not accurately and appropriately address the research question. Furthermore, since the coding and quantifying of the data have not been made explicit to the readers and the statistical procedures might have been misused, it is evident that the hypothesis has not been appropriately tested. Whilst the authors have included descriptive statistics of the results, which show that the participants have more responses in the VR condition than in the Reflection condition, they could not make any legitimate statistical inference that the two groups are significantly different.

On the basis of the manner in which the research has been constructed and presented, there is evidence to suggest, to a large extent, the conclusions drawn by the authors that “VR provides us with new contexts by letting us explore new peculiar perspectives and develop different skills” and that “we are induced to assume that such possibilities depend on the particular mental operations which are elicited by VR environments and which differ deeply from those occurring in traditional school settings” (Antonietti & Cantoia, 2000; p. 221) cannot be justified.

References

Anderson, J. R. (1995). *Cognitive Psychology and its Implication* (4th ed.). New York: Freeman.

Antonietti, A., & Cantoia, M. (2000). To see a painting versus to walk in a painting: An experiment on sense-making through virtual reality. *Computers and Education, 34*, 213-223.

Brown, A. J., & Dowling, P. C. (1998). *Doing research/ reading research: A mode of interrogation for education*. London: Falmer Press.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol Analysis: Verbal Reports as Data*. London: The MIT Press.

Ferrington, G., & Loge, K. (1992). Virtual reality: A new learning environment. *The Computing Teacher, April*, 16-19.

Robson, C. (1994). *Experiment, Design and Statistics in Psychology* (3rd ed.). London: Penguin.